

# HARVESTING

THE RESULTS AND IMPACT OF RESEARCH BASED  
ON HISTORICAL LONGITUDINAL DATABASES



**RADBOUD  
UNIVERSITY  
PRESS**

EDITED BY  
**Sören Edvinsson  
Kees Mandemakers  
Ken R. Smith  
Paul Puschmann**

Harvesting



# Harvesting

**The Results and Impact of Research based  
on Historical Longitudinal Databases**

Edited by Sören Edvinsson, Kees Mandemakers,  
Ken R. Smith & Paul Puschmann



## Harvesting

The Results and Impact of Research based on Historical Longitudinal Databases

Published by RADBOUD UNIVERSITY PRESS

Postbus 9100, 6500 HA Nijmegen, the Netherlands

[www.radbouduniversitypress.nl](http://www.radbouduniversitypress.nl) | [radbouduniversitypress@ru.nl](mailto:radbouduniversitypress@ru.nl)

Cover image: *Corn Harvest in Provence*, Vincent Van Gogh, 1888

Cover design: Textcetera

Lay-out: Marja Koster

Print and distribution: Pumbo.nl

ISBN: 978-94-9329-618-3

DOI: 10.54195/HYLR8777

Version: 2023-08

Download book via: [www.radbouduniversitypress.nl](http://www.radbouduniversitypress.nl)

Download individual chapters: <https://hlcs.nl/specialissue4>

© 2023, Sören Edvinsson, Kees Mandemakers, Ken R. Smith and Paul Puschmann

All authors hold copyrights of their own contribution.

**RADBOUD  
UNIVERSITY  
PRESS**

This is an Open Access book published under the terms of Creative Commons Attribution-Noncommercial-NoDerivatives International license (CC BY-NC-ND 4.0). This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, for non-commercial purposes only, and only so long as attribution is given to the creator, see <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

# Table of contents

Preface	7
Introduction: Major Databases with Historical Longitudinal Population Data: Development, Impact and Results <i>Sören Edvinsson, Kees Mandemakers, Ken R. Smith</i>	9
Dutch Lives. The Historical Sample of the Netherlands (1987–): Development and Research <i>Kees Mandemakers, Jan Kok</i>	15
Historical Chinese Microdata. 40 Years of Dataset Construction by the Lee-Campbell Research Group <i>Cameron Campbell, James Lee</i>	61
A Database for the Future. Major Contributions from 47 Years of Database Development and Research at the Demographic Data Base <i>Sören Edvinsson, Elisabeth Engberg</i>	89
The Utah Population Database. The Legacy of Four Decades of Demographic Research <i>Ken. R. Smith, Geraldine P. Mineau</i>	113
The Long Road to Health and Prosperity, Southern Sweden, 1765-2015. Research Contributions From the Scanian Economic-Demographic Database (SEDD) <i>Tommy Bengtsson, Martin Dribe</i>	139
The Impact of Microdata in Norwegian Historiography 1970 to 2020 <i>Hilde L. Sommereth, Gunnar Thorvaldsen</i>	163
Historical Life Courses and Family Reconstitutions. The Scientific Impact of the Antwerp COR*-Database <i>Paul Puschmann, Hideko Matsuo, Koen Matthijs</i>	187



# Preface

The past half century witnessed the birth of a multitude of large historical longitudinal population databases. These databases allow to describe and analyze the lives of our ancestors, the households they lived in, their families and social networks and their entanglement with the living environment. As a result, a tremendous number of studies on topics like marriage, fertility, divorce, social inequality, social mobility, migration, social inclusion, household formation, health and mortality have been conducted. The construction of the large historical longitudinal population databases has profoundly changed the way historians and social scientists conduct empirical research on past societies. With millions of records on the lives of our ancestors becoming available, researchers have moved from the analysis of aggregated cross-sectional data to historical life course analysis, allowing to test theories and hypotheses in the fields of (historical) demography, epidemiology, sociology, anthropology, and related fields, in a much more direct way, and to uncover causal patterns and pathways, ultimately leading to a deeper understanding of human behavior in the past.

In the past three years, my colleagues Sören Edvinsson (Umeå University), Kees Mandemakers (International Institute of Social History, Amsterdam & Erasmus University, Rotterdam) and Ken Smith (University of Utah) have edited a special issue entitled 'Major Databases with Historical Longitudinal Population Data. Development, Impact and Results' in *Historical Life Course Studies*. The collection brings together seven contributions that describe the results and impact of studies based on large historical population databases in several Western countries and China.

In 2016, on the occasion of the 2d conference of the European Society of Historical Demography, a book was distributed titled *The Future of Historical Demography* (Matthijs et al, 2016). The contributors sketched research vistas, interdisciplinary opportunities and promises of new sources and data. I feel this special issue shows how promises of 'big data' and innovative research methods have materialized in exciting results. Therefore, on the occasion of the 5th conference of the European Society of Historical Demography, we have turned this special issue into an edited volume by Radboud University Press.

This volume was made possible thanks to generous grants from Radboud University – i.e., the Faculty of Arts, the International Office Arts and the Radboud Group for Historical Demography and Family History, as well as HiDO, the international network of Historical Demography (Research Foundation Flanders), and the N.W. Posthumus Institute, the Research School for Economic and Social History in the Netherlands and Flanders.

Dr. Paul Puschmann

Assistant Professor of Economic, Social and Demographic History, Radboud Group for Historical Demography and Family History, Radboud University, Nijmegen

Co-editor-in-chief of *Historical Life Course Studies*, International Institute of Social History, Amsterdam

## Reference

Matthijs, K., Hin, S., Kok, J., & Matsuo, H. (Eds.). (2016). *The future of historical demography. Upside down and inside out*. Leuven/Den Haag: Acco. Available from <https://soc.kuleuven.be/ceso/fapos/publications/the-future-of-historical-demography-upside-down-and-inside-out>.



HISTORICAL LIFE COURSE STUDIES  
VOLUME 13 (2023), published 30-05-2023

# Introduction: Major Databases with Historical Longitudinal Population Data: Development, Impact and Results

Sören Edvinsson

Centre for Demographic and Ageing Research, Umeå University

Kees Mandemakers

International Institute of Social History, Amsterdam & Erasmus University Rotterdam

Ken R. Smith

University of Utah

## ABSTRACT

Over the last 60 years several major historical databases with reconstructed life courses of large populations spanning decades have been launched. The development of these databases is indicative of considerable investments that have greatly expanded the possibilities for new research within the fields of history, demography, sociology, as well as other disciplines. In this volume spanning seven articles, eight databases are included that have had a wide impact on research in various disciplines. Each database had its own unique genesis that is well described in the articles assembled in this volume. They inform readers about how these databases have changed the course of research in historical demography and related disciplines, how settled findings were challenged or confirmed, and how innovative investigations were launched and implemented. In the end we explore how research with this kind of databases will develop in future.

**Keywords:** Historical demography, Historical microdata, Life course, Social science history, Record linkage, Longitudinal research

PID article: <https://doi.org/10.51964/hlcs14840>

The article can be downloaded from [here](#).

© 2023, Edvinsson, Mandemakers, Smith

This open-access work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/), which permits use, reproduction & distribution in any medium for non-commercial purposes, provided the original author(s) and source are given credit. See <http://creativecommons.org/licenses/>.

## 1 AIMS AND CONTENT

Over the last 60 years several major historical databases with reconstructed life courses of large populations have been launched. The development of these databases is indicative of considerable investments that have greatly expanded the possibilities for new research within the fields of history, demography, sociology, as well as other disciplines. At the annual meeting of the Social Science History Association in Montréal in 2017, the session "Development of Major Databases and Their Results From the Beginning Till now" brought together presentations from some of the largest and most well-established databases with life course data, databases that have also been at the forefront of the development in this field. We were well aware in 2017 of numerous additional databases that had been established around the world in recent decades. In his valedictory speech Kees Mandemakers (2023) made an inventory of a total of 54 databases and even this compilation is not exhaustive.

In order to collect, organize, and then publish information on these major databases in a single collection, invitations were first sent to the leaders of about 25 of these databases. We received in most cases positive and enthusiastic reactions and, in case the leaders of a database declined cooperation, it was mostly due to time constraints. We had no specific selection criteria, except that databases had to be actively used and maintained and the primary purpose of the database had to be the (re)construction of individual-level historical life courses. Archived databases, like the Louis Henry dataset (Séguy, 2001) were therefore excluded. Following the first round of invitations, still others joined the collective endeavour, expanding the geographic coverage of our collection. All in all, we are happy to have the assembled contributions representing 24 databases in two special volumes of *Historical Life Course Studies*. The number and diversity of databases represented here is truly impressive!

Our overall strategy of describing these major databases resulted in creating two separate special issues. One, *Content, Design and Structure of Major Databases with Historical Longitudinal Population Data*, edited by George Alter, Kees Mandemakers and Hélène Vézina, deals with the technical and organizational aspects of these databases. They concentrate on aspects such as their origins and evolution including any setbacks, dependence on external funding, content and database designs. The present volume focuses on how the databases contributed to discoveries and insights in historical demography and related fields. Several questions were posed to the leaders of each contributing database: How were previous research questions addressed or altered? What novel lines of inquiry were developed, thanks to the availability of your data? What new knowledge, insights, and scholarly debates were generated that were attributed to the availability of your data? The objective of these central questions was to explore the research productivity and impact of these databases. To summarize the outcome of the investments and labor in a comprehensive way, each contributor was asked to describe the main research contributions resulting from the use of their data, especially with regards to knowledge and insights, that were (1) unavailable before the databases were constructed and (2) could not have been accomplished without having access to individual life course data. In general, the articles assembled in this volume inform readers about how these databases have changed the course of research in historical demography and related disciplines, how settled findings were challenged or confirmed, and how innovative investigations were launched and implemented.

In this volume spanning seven articles, eight databases are included that had a wide impact on research in various disciplines. For databases that are still in early stages of development or for other reasons had limited impact until now, we decided to include information about their impact in the technical articles in the volume mentioned above. This means that the databases included in this impact volume all have their counterpart in the technical one. There is always an exception: the Historical Sample of the Netherlands (HSN), on which the first published article in our volume appeared, does not have a technical counterpart, due to a lack of time. On the other hand, one of the described offshoots of the HSN, the LINKS database, found its own place in the technical volume. The other six databases describing their impact are the Historical Chinese Micro Database, the Demographic Database of Umeå, the Utah Population Database, the Scanian Economic Demographic Database of Lund (SEDD), the Norwegian Historical Micro Database and the Antwerp COR\*- Database. Unfortunately, two major databases that presented a technical article had insufficient time to deliver an impact contribution. These are the two representing Quebec: BALSAC in Chicoutimi and the Programme de Recherche en Démographie Historique (PRDH) in Montreal.

## 2 CONTEXT AND TYPOLOGY OF THE DATABASES

Databases with life course data may be distinguished into databases containing (I) longitudinal data, (II) family reconstitutions and (III) semi-longitudinal data. The differences between I, II and III, is that the latter two categories contain data whereby persons are not reconstructed in a continuous way but on the basis of linked sources, e.g., censuses or, in case of family reconstitutions, only church or civil records. Half of our included databases, HSN, SEDD, DDB Umeå and COR, belong to the 'pure' longitudinal datasets; Utah, Norway and China are of a semi-longitudinal nature; and the LINKS database can be categorized as a family reconstitution dataset.

Each database represented in this volume has its own unique genesis that is well described in the various papers. For example, the launch of the DDB at Umeå University was initially motivated by an interest in the development of literacy. For the Utah Population Database, the impetus was the focus on genealogies, genetics, medicine and family history. At the same time, several common elements and circumstances connect these distinct databases. Their developmental arcs share, for instance, commonalities. Perhaps foremost, most of the databases stand on the shoulders of giants who championed quantitative history and the history of the ordinary person. This includes members of the Cambridge Group for the History of Population and Social Structure (Wrigley, Davies, Oeppen & Schofield, 1997), the Annales School with its advocacy of social history (Séguy, 2016), and the proponents of the life course perspective arguing for the plasticity of human development and the role of history (Kok, 2007). With these intellectual foundations as bedrock, the inevitable influence of technology proved to be a catalyst for accelerating the insights of quantitative history by digitizing archival records and through record linking methodologies that served to reveal the complexities endemic in the reconstruction of human populations.

The birth dates of the distinct databases vary but three in our volume originate from the 1970's when computers and software facilitated data entry, processing and database management. These are the Demographic Database Umeå, the Utah Population Database and the Norwegian Historical Data Centre in Tromsø. Others were the *Registre de la population du Québec ancien* (Université de Montréal) and the BALSAC database (Université du Québec à Chicoutimi). These efforts were followed in the early 1990s with new databases here represented by the SEDD database in Sweden, the Chinese datasets and the HSN database in the Netherlands. More recent work has led to the launch of new databases of which we include the LINKS database and the Antwerp COR\*-database. But most recent databases were actually launched in other parts of the world, including Asia, Australia, and South Africa. Many are described in the technical volume, including a summary of their impact. While the expansion of these infrastructures is impressive and benefits the research community broadly, there remain significant portions of the globe that are not represented, largely due to the lack of archival data and a lack of resources needed to create and maintain complex databases.

Accordingly, of the databases represented in this special issue, the majority are derived from western Europe and North America. Many of these were launched decades ago and have created a legacy through numerous publications, large numbers of trainees, and the development of reliable infrastructures that speak to their stability. From the body of work derived from these databases represented in this volume, we discuss some recurrent themes on the way databases have contributed to the literature, have offered new findings, especially results that would not have been possible without these types of databases. In this overview, we have also introduced some information on findings from the other special issue on the technical aspects of the databases. There we also find more representation of databases from other parts of the world including Korea, Japan, South Africa, Australia, Tasmania, Russia and Spain.

## 3 IMPACT ON RESEARCH

For much of the western world, basic demographic trends and structures were well described and analyzed before the large individual and family level life course databases were developed. But it is worth emphasizing that this is not the case everywhere. More important is that these classic data sources, often highly aggregated, could not address a range of questions about demographic and social mechanisms, such as the role that historic shocks have on individual life course trajectories. To analyze these issues data needs to be collected on a micro level. The databases represented in this

volume have both corroborated past findings while also expanding our understanding in new ways given the richness of the databases and the tools used to analyze them. We will highlight here some of the results for some important aspects of social and demographic history.

*Subpopulations:* While the databases attempt to characterize entire populations in a geographic area, thereby allowing for analyses of all individuals and families, it is clear that the large "sample size" of these databases allows for examining demographic heterogeneity present in all populations. Comparative analyses of the demography for different groups are a common aspect of all the databases represented. A recurrent theme in nearly all the papers relate to differences in racial or ethnic groups, religion, sex, and age.

*History Meets Lives:* A significant advantage of the databases with individual information is that it makes it possible to study how individuals respond to external circumstances, and societal structures and their changes. Shocks may be localized, allowing for comparisons between regions and over time. If you view the dynamics of entire populations as representing what history's video camera has captured (and therefore visible in a database), it is possible to examine intensively and extensively how history shapes lives. Indeed, when intentional changes happen through policy implementation, evaluations of those policies have been conducted.

*Life Course of Individuals and Families:* One of the most important aspects of research using these databases is their life course coverage often from birth to death. A multitude of studies have analyzed different aspects of this, something that was not possible before the development of these databases that cover entire life-spans. This is the case both at individual as well as the family and household level. These dynamics in families and households have been extensively studied as noted in this volume.

*Intergenerational and Familial Studies:* Given the time scale encompassed in these databases, it is possible to see events comprising entire lives for individuals and their ancestors and descendants. As long as included individuals remain in the catchment area covered by the database, it is possible to observe connections, both genetic and social in origin, among relatives as well as those who marry into a lineage. Combining these linkages with information about historical events provides opportunities to see the potential intersection of social history and family history. These opportunities have attracted not only historical demographers but also geneticists and evolutionary biologists.

*Comparative Analyses:* The growing number of large historical databases available to the research community create the opportunity for examining common research questions across social and historical contexts. Given the heterogeneity of decades and locations embedded in these historical databases, it is now possible more than ever to expand on comparative analyses that involve a larger number of populations, all of which serve to improve our understanding of social and family history.

*Database Expansion:* Given the success and productivity of the large historical databases included here, it is reasonable to consider adding to this portfolio by building new resources from previously omitted areas of the world. Offering expertise and direction to investigators who represent previously underdeveloped historical populations from those who have already done it can prove to be an effective way to improve data coverage.

## 4 FUTURE DIRECTIONS

As we have introduced this volume on history and demography and key databases, we have likewise adopted a more historical perspective about their intellectual and scientific impact — that is, what these databases have done. Here we briefly consider what these databases could do going forward. The existing databases saw their births in a time where the central records were decidedly amenable to the social sciences. This meant that archival, census and religious sources and the demographic, family, and spatial information they contained formed the essential ingredients needed in the database recipe. With this as foundation, we are now imagining how to build off of it. Several new opportunities now (and in some cases have been) present themselves. First, the advent of population genetics means that the growing number of countries that collect genetic material may be used to link to historical databases, assuming the appropriate human subject protections are in place. One can imagine more

collaborations between historical demographers and geneticists studying disease risk as well as demographic phenomena like fertility and mortality using DNA information. Indeed, knowing genetic variants in contemporary populations, when joined with family data in the databases, may allow one to infer what these variants are in past populations (Adams, Lam, Hermalin, & Smouse, 1990).

Second, the rise of geographic information systems (GIS) has paralleled the rise of big genetics and in like fashion, offer new directions that build on the existing historical databases. Since the databases described in this volume generally have geographic information within them, it is now possible to link, at a spatial level, additional data to individuals and families. An excellent example of this are the data available in the IPUMS National Historical Geographic Information System (NHGIS) in the US. But in general, joining these GIS variables can only enrich the research opportunities that will provide novel insights into our understanding of historical populations.

Finally, it is possible to consider how to do comparative analyses in new ways. Specifically, the comparison can now be done by actually connecting people and lineages in one population to another. In the spirit of linking European and North American databases, it is possible to observe and analyze people who left one country for another in these databases and compare them to people who did not emigrate. So, in addition to studying two independent populations, it is conceivable to see the same person or their relatives and descendants in both countries.

## REFERENCES

- Adams, J., Lam, D. A., Hermalin, A. I. & Smouse, P. E. (1990). *Convergent issues in genetics and demography*. New York: Oxford University Press.
- Kok, J. (2007). Principles and prospects of the life course paradigm. *Annales de démographie historique*, 113(1), 203–230. doi: [10.3917/adh.113.0203](https://doi.org/10.3917/adh.113.0203)
- Mandemakers, K. (2023, January 20). “You really got me”. *Ontwikkeling en toekomst van historische databestanden met microdata* [Development and future of historical databases with microdata] (Valedictory speech). Erasmus University, Rotterdam, the Netherlands. doi: [10.25397/eur.23256467.v1](https://doi.org/10.25397/eur.23256467.v1)
- Séguy, I. (2001). *La population de la France de 1670 à 1829: l'Enquête Louis Henry et ses données* [The population of France from 1670 to 1829: The Louis Henry survey and its data]. Paris: INED.
- Séguy, I. (2016). The French school of historical demography (1950–2000). In: A. Fauve-Chamoux, I. Bolovan, & S. Sogner (Eds.). *A global history of historical demography. Half a century of interdisciplinarity* (pp. 257–276). Bern: Peter Lang.
- Wrigley, E. A., Davies, R. S., Oeppen, J. E., & Schofield, R.S. (1997). *English population history from family reconstitution 1580–1837*. Cambridge: Cambridge University Press. doi: [10.1017/CBO9780511660344](https://doi.org/10.1017/CBO9780511660344)



HISTORICAL LIFE COURSE STUDIES  
VOLUME 9 (2020), published 15-06-2020

## Dutch Lives

### The Historical Sample of the Netherlands (1987–): Development and Research

Kees Mandemakers

International Institute of Social History, Amsterdam & Erasmus University Rotterdam

Jan Kok

Radboud University Nijmegen

#### ABSTRACT

The HSN was initiated during the period 1987–1989 when an interdisciplinary and interuniversity group of Dutch scholars started discussing the foundation of one large database with data on individuals. Building one general prospective database with multiple research possibilities was considered as the only way to realize a cost-effective and properly documented tool for historical research from economic, social, demographic, epidemiological and geographic perspective. The birth registration was considered the most adequate sample framework. The new database should be 'open' in the sense that extension should be possible in all kinds of ways: more sources or variables, more persons and larger time periods. The HSN was deliberately created as a nationwide sample covering the whole 19th and 20th century. Since 1991 about 12 million Euro has been invested in the database and related projects. Besides the basic sample about 25 additional projects have been realized that created all kind of extensions to the database. A special project is LINKS by which the indices of names from the Dutch civil registration are used to reconstruct pedigrees (for the period 1780–1940) and complete families (1811–1900) for the whole of the Netherlands or parts of it. In this article we will present an overview of the research that was done with the original themes and the new fields that were introduced over the years. We will also go into methodological issues that were picked up by the 'HSN community' and we will point out the present and future challenges for the HSN.

**Keywords:** Historical databases, Life courses, Demography, Sociology, Epidemiology, History, Economy

DOI article: <https://doi.org/10.51964/hlcs9298>

© 2020, Mandemakers, Kok

This open-access work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/), which permits use, reproduction & distribution in any medium for non-commercial purposes, provided the original author(s) and source are given credit. See <http://creativecommons.org/licenses/>.

## 1 INTRODUCTION

In their recent overview of the history of historical demography in the Netherlands, Theo Engelen and Ad van der Woude† (2016) describe the years around 1990 as a 'turning point', a period in which Dutch historical demographers began to move away from aggregated data based on censuses to individual-level data. More than class, group, local community or region, the individual was taken as the center of research. In this development, the Historical Sample of the Netherlands (HSN) is seen as 'the most prominent example'.

The HSN was initiated during the period 1987–1989 when an interdisciplinary and interuniversity group of Dutch scholars started discussing the foundation of one large database with data on individuals to be used in existing research and for exploring new research themes. Starting a database with data on the appropriate level of research was indeed the main motivation behind the formation of this group of scholars. Out of necessity much research was using data on the aggregative level of provinces or municipalities while decisions regarding life course issues were taken on an individual level. Building one general database with multiple research possibilities was considered as the only way to realize a cost-effective and properly documented tool for historical research from economic, social, demographic, epidemiological and geographic perspectives. During this initial period some important decisions were taken. The sample should be taken from an individual perspective nationwide and be limited to the 19th and the early 20th centuries. This creates standardized biographies for the whole 19th and 20th century, even until the present day. This choice for a prospective database also implied that birth certificates were the best documents to start with and to form the sample framework. The new database should be 'open' in the sense that extension should be possible in all kinds of ways: more sources or variables, more persons and larger time periods. Besides the original research persons, it should also be open to include additional research persons (such as siblings or a second and third generation) and to make oversampling from birth certificates for some specific region or time period. Other advantages of the database would be the systematization in data gathering and documentation, the reuse and multiple use of existing data, economies of scale by working on a national level and the possibilities to contextualize more in-depth regional research in a national context (Mandemakers, 1989, p. 96). The HSN was not the first database in this field; other databases already existed, in particular in Sweden and Québec (Hall, McCaa, & Thorvaldsen, 2000). The Dutch example distinguished itself by being a sample (compared with Sweden) working on a national instead of regional scale (see also the HSN website: <https://iisg.amsterdam/en/hsn>).<sup>1</sup>

In 1991, the HSN found its home at the International Institute of Social History (IISH) in Amsterdam and a pilot project started with a sample from birth certificates in the province of Utrecht for the birth period 1812–1922. In 1993, the HSN started collecting data from population registers, as well. In the beginning this was done only for additional projects and on a limited scale. Since 1991 about 12 million Euro has been invested in the database and related projects: 3 million by the IISH and 9 million by way of external funding, especially the Dutch Research Council (NWO). Besides the basic sample, about 25 additional projects have been realized. This has created all kind of extensions to the database. A special project, LINKS, uses the indices of names from the Dutch civil registration to reconstruct pedigrees (for the period 1780–1940) and complete families (1811–1900) for the whole of the Netherlands or parts of it. The HSN and the additional projects have been widely used in hundreds of presentations and scientific publications. By March 2019 there has been about 400 scientific publications and 15 dissertations completed wholly or partly based on HSN or LINKS data. The first results of these studies were communicated in over 700 worldwide presentations.<sup>2</sup> The large output can be linked to the fact that the data has always been freely available.<sup>3</sup> In a quantitative sense, HSN can certainly be considered to have been a game-changer in the fields of Dutch historical science, especially historical demography and historical sociology. But what has been its impact in a *qualitative* sense? With HSN,

1 The first overview was given by Hall, McCaa, and Thorvaldsen (2000). For the most comprehensive overview see the website of the European Historical Population Samples network: <https://ehps-net.eu/databases>.

2 For the latest overview, see <https://iisg.amsterdam/en/hsn/products/publications> and <https://iisg.amsterdam/en/hsn/products/presentations>.

3 The data of both HSN and LINKS are freely available for scientific research after signing a license agreement. License forms can be requested by writing an email to [hsn@iisg.nl](mailto:hsn@iisg.nl). By signing a license the researcher guarantees the private character of the data, to use the data only for science in a non-commercial way and to share the results of his/her research with the HSN, for more information, see <https://iisg.amsterdam/en/hsn/privacy-statement>; <https://iisg.amsterdam/en/hsn/products>.

Dutch social science historians have added a new perspective to Dutch history. But they also analyzed the Dutch experience in the context of more international debates on 'universal' behavioral patterns. Regularly, this contribution took the form of comparative projects.

We especially focus on debates: to what (old) questions has HSN provided the answers? Have definite outcomes been reached or is research still going on? In this article we will present an overview of the research that was done on the original themes and the new fields that were introduced over the years. We will also go into methodological issues that were picked up by the 'HSN community' and which helped the field forwards. In the last section we will point out the present and future challenges for the HSN, both in terms of methods and data.

## 2 STRATEGY AND DATABASE DEVELOPMENT OF THE HSN

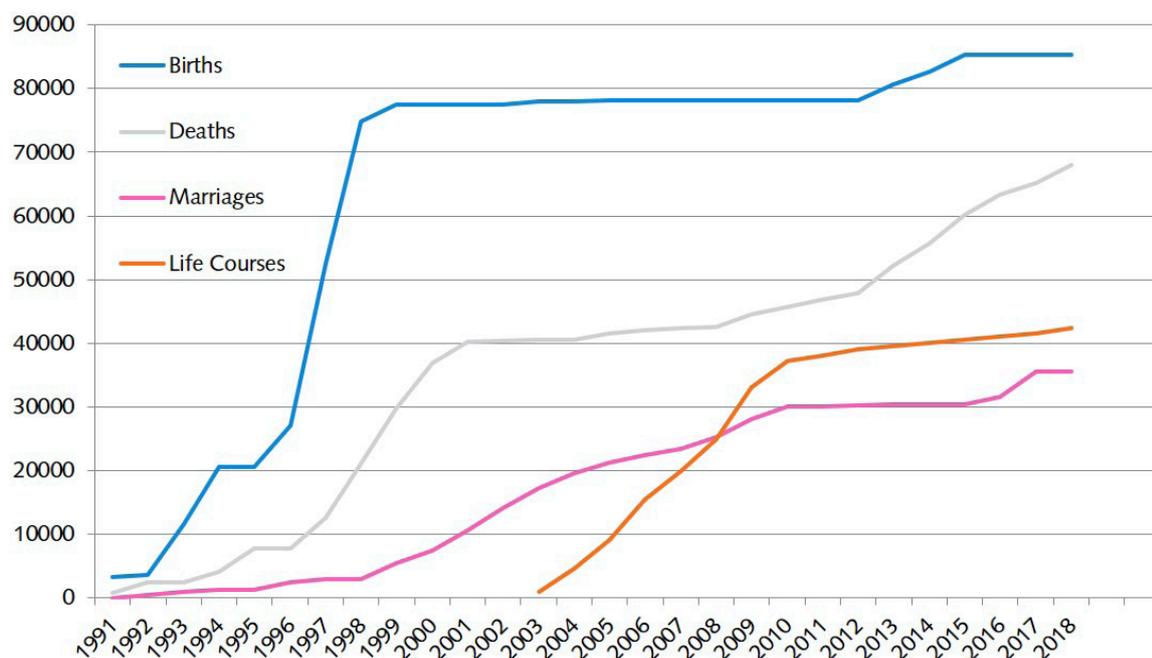
Before answering the main question, we will present a more detailed overview of the development of the HSN database. The incremental growth of the database implies that earlier studies were based on different releases (in terms of period, region and sources) than later ones. Thus, the development of HSN needs to be explained to understand shifts in research questions. What did the initiators actually have in mind when starting the HSN and were their ambitions realized? The HSN operates in a dynamic setting. Firstly, the scientific fields in which the HSN operates are constantly changing in focus and methods. Over and over again new questions have to be answered, leading to the need for extensions of the HSN dataset, especially by way of additional projects, resulting in extra samples and/or including data from new sources. Secondly, the HSN itself plays an important role in agenda setting and the dynamics of research, especially in the Netherlands. An example is the introduction of the integrating concept of the 'life course' in Dutch historical research for which the data collected by the HSN were extremely useful (Kok, 2000). Thirdly, the possibilities for research were also determined by the logic in the construction of the database itself in coherence with the way investment grants were allocated. Fourthly, the HSN became an international player in the development and documentation of best practices in the field of historical life course databases.

The database activities of the HSN can be distinguished in three categories: a) projects that are focused on the construction of the central database which is the HSN basic sample with life courses, b) projects that are initiated by researchers or institutions adding new research persons to the database (for example siblings or children of existing research persons) and/or adding data that are not considered to be part of the central database such as cadastral or tax data and c) projects to develop software for data entry and data integration and to document the expertise by which all projects are accomplished, archived and released for research (Mandemakers, 2001a, 2006a). The focus of the central database is the systematic collection of data from civil certificates (birth, marriage and death) and the population register.<sup>4</sup> The research persons (RPs) of the HSN are selected by way of a stratified random sample from the birth certificates of the period 1812–1922. The strata are defined by ten yearly time periods and by region. The regional stratification was based on the administrative division in provinces and the provinces were further stratified in urban and countryside areas. The province of North-Holland for example was stratified in four parts: Amsterdam, large cities (Haarlem, Alkmaar, Hoorn and Zaandam), small cities and the countryside. Eleven time periods and 25 regional areas made 275 strata all together, making sure that the sample of the HSN would be representative in time and in geographical perspective. This representativeness was further guaranteed by varying the sample fraction from 0.75% for the period 1812–1872 to 0.5% for the period 1873–1922 (Mandemakers, 2000, pp. 151–155). The difference between these two levels was motivated by the wish to get a more or less equal absolute number of survivors at the age of 16 for each ten yearly cohort, depending on the number of births and the infant and child mortality which varied between 15 and 40% for the period till 1870 (Engelen, 2009; van Poppel & Mandemakers, 2002). The number of sampled persons for the central database amounts to about 85,500 persons. This is about 0.6% of the 14.5 million persons that were born in this period in the Netherlands. The goal of the HSN is the reconstruction of the life courses of all these sampled persons.

4 For a description of the variables that can be collected from these sources, see the website of the HSN (<https://iisg.amsterdam/en/hsn/data/sources>) or Mandemakers (2006b), Vulsmas (2002).

Figure 1 shows the development of the construction of the HSN central database. In 1991 the HSN started with a pilot project in the province of Utrecht that did not only focus on entering birth certificates but was also directed at the data entry of death certificates of persons who died before the age of ten, of personal cards administered for all persons living in the Netherlands from the 1st of January 1940 onwards and of marriage certificates. In 1993 the pilot project ended successfully (Mandemakers & Boonstra, 1995) and new funding was received to continue in the same way with the provinces of Zeeland and Zuid-Holland. By way of grants from investment funds of the Dutch Research Council (NWO) the HSN was scaled up to the nationwide level between 1996 and 2002. During this period most of the sampled persons were entered into the database. Only the period 1903–1922 lagged behind, since initially for this period a sample of 0.25% was constructed. Getting a sample with a comparable number of surviving 16-year-olds for each time period, on average 0.35% to 0.4% would be the norm for this period. For reasons of comparability and because three provinces had already been sampled at the 0.5% level, it was decided to bring the whole period to the same level of 0.5%. Data entry of death and marriage certificates was limited to these certificates that were easy to find, searching only in the municipality of birth and neighboring municipalities and in case of the death certificates only for those who died at a young age. From 2010 onwards it became much easier to find death and marriage certificates because of the digital indices that were created by the Dutch regional archives mobilizing thousands of volunteers. And nowadays more and more persons can be traced in the scans of all civil certificates made available by the municipal and regional archives (see <https://www.wiewaswie.nl/en>). This means that deaths outside the municipality of birth are getting more and more complete, especially for the births after 1862 when all life courses have been followed by way of population registers.

Figure 1 Development and size of the HSN central database, 1991–2018

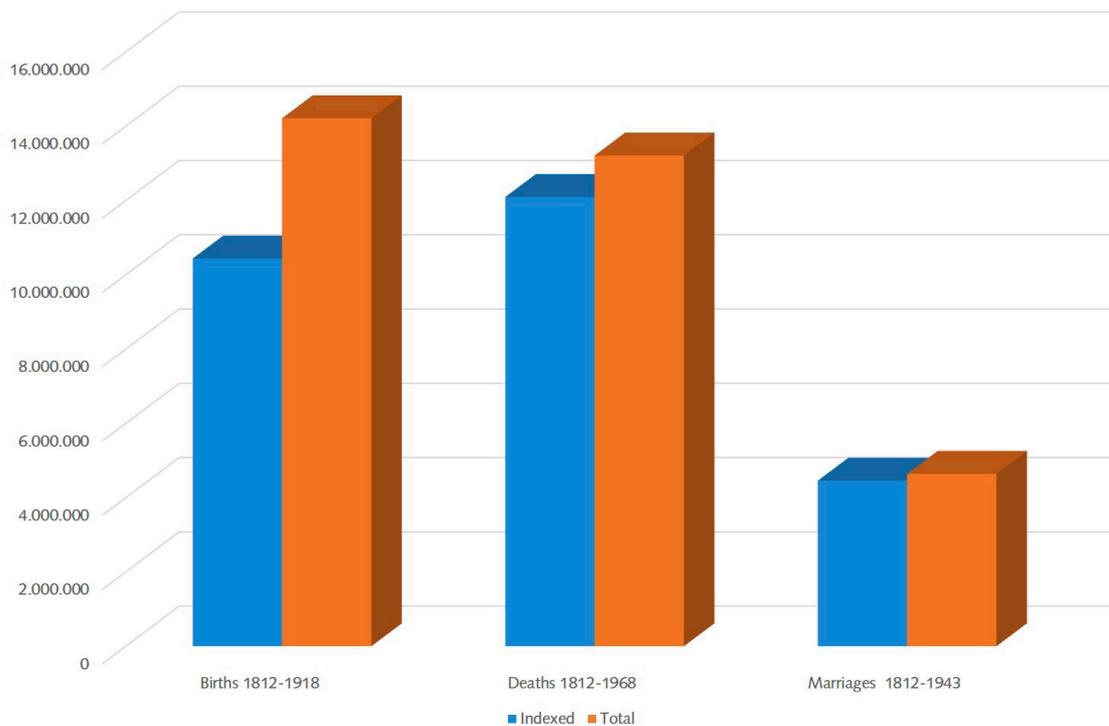


Soon, researchers showed up expressing the wish to collect data from population registers. In a limited way this was started up in 1993 in the form of three additional projects: *Migration in the province of Utrecht* (Jan Kok), the epidemiological project *Reduced fecundity because of maternal high-risk conceptions* (Luc Smits, Gerard Zielhuis and Piet Jongbloet) and *Regional differences in demographic behaviour, the Netherlands, 1900–1960* (Angélique Janssens). The system of population registers was introduced in the Netherlands in 1850, following the Belgian lead. Basically, the registers updated census information on a day-to-day basis: this means that every change in household composition (through birth, death, marriage, arrival, or departure) was to be recorded with a proper date. The registers also provide information such as place and date of birth, religion, occupation and (family) relation to the head of the household. As people can be traced in all their households across the country, this source allows for a complete and very detailed life course reconstruction of RPs born after 1850 (Mandemakers, 2006b).

With a great investment impulse from NWO, for the project *Life Courses in Context* (2003–2008), it became possible to collect all population registers for the then sampled persons from the birth period 1863–1922 in a systematic way (Mandemakers, 2004). The outcomes of this project also provided sufficient information to collect death and marriage certificates for these persons. Priority was given to the sampled persons from the provinces Utrecht, Friesland, Zeeland and Rotterdam. This priority was meant to quickly offer researchers workable datasets, since for these regions a lot of data from the registers had already been collected within the framework of the project *Early-life conditions, Social mobility and Longevity* (George Alter and Frans van Poppel).<sup>5</sup> This also implied that for this group the period of birth was extended to 1850–1862. Besides the RP itself, other relatives are included in the database as long as they lived together in a family or household context with the RP. For an average RP who reached the age of 60 years we count about 26 other persons (2 parents, 2 parents-in-law, 1 spouse, 8 siblings, 6 children, 6 witnesses from the certificates and 2 other kinds of kinship).

Since the early 1990s, more than a thousand volunteers have digitized all names, ages and partly also occupational titles and birth places from birth, marriage and death certificates as far as they are open to the public. Initially the database was known as GENLIAS, now it is called WieWasWie ('WhoWasWho', abbreviated as WWW) and presently it includes over 200 million persons' appearances. All digitized data are publicly accessible through the website of WieWasWie ('WhoWasWho', <https://www.wiewaswie.nl/en>), based at the Dutch Center for Family History (CBG).<sup>6</sup>

Figure 2 Total and indexed civil certificates, Netherlands, September 2018



The main part of WWW consists of indices from civil certificates. For the whole of the Netherlands the civil registration system started in 1811 and the registers are public with a delay of 100, 75 or 50 years depending on the type of certificate. At present, the registers of birth until 1919, the marriage registers until 1944 and the registers of death until 1969 are public. In the middle of 2018, about 26 million civil certificates have been digitized, containing information on about 110 million appearances of (not unique) persons. As figure 2 shows the marriage registers have been indexed quite completely, lagging behind are the birth and to a lesser extent the death certificates, especially those of Amsterdam. Basically, the system includes the full names of the main actors in these certificates: the birth, the

5 For more information about all additional projects, see <https://iisg.amsterdam/en/hsn/projects> and the annual reports of the HSN (<https://iisg.amsterdam/en/hsn/annual-reports>).

6 CBG is the national center for knowledge, documentation and publicity for genealogy and heraldry. It is a non-commercial organization, supported by about 40,000 contributors. The work of the volunteers is managed by more than twenty cooperating regional archives.

death, the bride and the groom and their parents. Since 2005 the HSN has distributed data from this system. With the LINKS program a more systematic approach was started in 2010.<sup>7</sup> LINKS stands for *LINKing System for historical family reconstruction*. By way of *record linkage software* certificates are being linked to create pedigrees and families. This matching is relatively easy, because the parents are always included in the data from the WhoWasWho system and because the original names of females are always kept in Dutch certificates. Other included variables are the age of the death and of the bride and groom and for some provinces the occupational titles of all actors (about 60% part of the data from the marriage records).<sup>8</sup> After all certificates have been entered, it will encompass the reconstruction of *all* 19th and early 20th century families in the Netherlands from 1840 onwards.

The products of the LINKS software system are delivered in specific releases, but the dataset as a whole is known as the LINKS database. And although these data must be seen as a separate database, it is very much connected to the HSN, in the sense of creation, content and dissemination, so we will integrate the results from LINKS with the results from the HSN in the following chapters. Based on birth certificates, all HSN Research Persons and its family members are included in this kinship system as well. This opens up enormous possibilities to enrich the existing life histories with kinship in the second, third and fourth degree. The first release with an integration of the HSN Life Courses and the LINKS database is expected for 2020. A systematic comparison of both HSN and LINKS has been made by van den Berg, van Dijk, Mourits, Slagboom, Janssens, and Mandemakers (2020). On the basis of a sample of an integrated dataset from the province of Zeeland they compared the average number of children within families, the average age at death for RPs and their kin and other demographic indicators. They concluded that RPs own families were reconstructed very thoroughly in HSN, which is not biased by out-migration as is the case with LINKS. However, information on parental families is more complete in LINKS.

This overview is limited to the basic concepts of both HSN and LINKS. Two separate articles are prepared to be published in this journal. In these articles both databases will be described from a more technical point of view. For the HSN this will include an overview of all projects that have been undertaken since the beginning including a description of all datasets that have been created as oversamplings from the main one (Mandemakers, forthcoming). For LINKS we will describe in more detail the way the software system works and the amount of data that are available (Mandemakers, Bloothoof, & Laan, forthcoming). Till now the HSN released about 70 data releases, often in relationship with a specific project. The first release of LINKS data appeared in 2005, since then over 30 data releases have been made available.<sup>9</sup>

## 3 NEW ANSWERS TO CLASSIC RESEARCH QUESTIONS

### 3.1 THE AMBITIONS OF THE HSN

In 1989, when the plans for the HSN were publicly announced, five research fields were explicitly mentioned that would profit on what was also called the 'birth-bank' (Mandemakers, 1989). The following examples were given:

- 1 Social stratification
- 2 Composition of the workforce
- 3 Social mobility
- 4 Migration
- 5 Demographic themes (mortality, marriage frequency, marriage age, marital fertility and bridal pregnancies)

7 LINKS was granted through the NWO CATCH program, see [https://www.nwo.nl/en/research-and-results/programmes/Continuous+Access+To+Cultural+Heritage+\(CATCH\)](https://www.nwo.nl/en/research-and-results/programmes/Continuous+Access+To+Cultural+Heritage+(CATCH)) and <https://iisg.amsterdam/en/hsn/projects/links>

8 For the death and especially the birth certificates the percentage of included occupational titles is much lower.

9 For the most recent overview of HSN and LINKS releases, see <https://iisg.amsterdam/en/hsn/products/releases> and <https://iisg.amsterdam/en/hsn/projects/links/links-releases>.

Although we know that these five themes were not meant to be exhaustive but were only presented as examples of possible research fields, we are still surprised by the original modest expectations compared to the underlying vision of the plans. The first two themes are quite related and we will band them together under the umbrella of social stratification. The HSN proved fruitful in getting a grip on the old theme of social mobility as it enabled Dutch researchers to build on the paradigm change by Blau and Duncan (1967) redirecting the focus from social mobility tables to an approach with status achievement models based on micro data (Heath, 1981). Classic debates in historical demography concern social inequality in relation with death, household composition, marriage (and its 'Malthusian' connection to survival and economic independence), and, above all, the demographic transition (Coale & Watkins, 1986). From the onset, HSN researchers used the new data to contribute to these debates. Together with social mobility they form the nucleus of this section.

The migration theme we consider as quite new since in the beginning this was only envisaged in a rather limited way by combining civil certificates (birth, marriage and death). This supplied a rough indication of migration histories. During the development of the database the use of the population register for migration studies became a very innovative aspect of the HSN. Since section 4 focuses on new types of research, we will handle migration there. But, of course the difference between new and old research themes is somewhat arbitrary.

## 3.2 LIFE CHANCES AND SOCIAL BACKGROUND

We will first briefly discuss the theme of social stratification. This topic was quickly superseded by research interest in the process of social mobility. The changing society as such did not become central in research rather it was chances of an individual to proceed into another social layer than the one he or she was born in. More and more social mobility was considered a process of status attainment. Research into processes of social background and mobility is interwoven with all kinds of other topics such as migration, nuptiality and mortality. So, also within the HSN community, social mobility is studied from all point of views. For the sake of argument, we have made a distinction into different sections: marriage mobility, status attainment, intragenerational or career mobility, educational mobility and female-specific aspects of mobility. However, since part of the mobility studies are overlapping with studies in the field of demography, some of the studies will not be discussed in this section but in the one on demography.

### 3.2.1 SOCIAL STRATIFICATION

The first theme focused on the solution of what became known as the 'great question of the social stratification' of the Netherlands between 1850 and 1940. The discussion was started by Giele and van Oenen (1974) and circled around two questions: should a social stratification be constructed from the perspective of social class or of social status, and subsequently how did this social structure develop (Boonstra & Mandemakers, 1995)? For all research persons born from 1812 until 1872 it would be possible to reconstruct a complete census for each moment in time until the age of 59 from 1872 onwards, for those born from 1812 until 1882 until the age of 69 from 1873 onwards, etc. However, until now the HSN-database has been developed mostly for persons born after 1863, so it is not surprising that the HSN has not been able to revitalize this debate. Anyway, not much new work has been published either. Exceptions are articles by Mandemakers (2001b) covering the period 1850–1990 and van Leeuwen and Maas (2007a) on economic specialization in which only the last authors made use of the HSN-database. Besides, the introduction of micro-data made the focus shift from changing class structures to individual chances of upward or downward social mobility.

Related to the theme of social stratification is the composition of the workforce. A problem here was the lack of the occupational censuses of 1869 and 1879. The 1869 census did not take place at all and the results of 1879 were only used as a mechanism to check the locations of residence (van Maarseveen, 2002). Again, as in the case of social stratification, the HSN was expected to be able to make a simulation of the census, and for the same reason (no life courses from the birth period before 1863) it did not work out. However, the HSN dataset was used to gain more knowledge on all kinds of aspects of the labor market, such as the position of migrants (Delger, 2003) or the occupations of women. The occupational titles on marriage certificates of the HSN were especially suitable to trace changes in registered female labor (van Poppel, van Dalen, & Walhout, 2009); combined with certificates of the LINKS database (Maas & van Leeuwen, 2006) and for a sample of twelve municipalities (Walhout & van Poppel, 2003).

### 3.2.2 INTERGENERATIONAL AND MARRIAGE MOBILITY

Intergenerational mobility has mainly been studied by using data from the marriage certificates which include the occupational title of both bride and groom and all four parents. Marriage mobility is a good indicator for the 'openness' of societies, and researchers expect that status and class differences in partner choice have diminished after 1850. This supposedly resulted from the increasing importance of 'romantic love', and from growing possibilities to reach a better social position than their parents, thanks to education and their own achievements (Kok & van Leeuwen, 2005).

Van Leeuwen and Maas (1995, 1997) started their research on the openness of society with the data of the Utrecht pilot project. They concluded that although there had been an increase in absolute social mobility this was a consequence of the growing opportunities in society while the relative mobility (mobility controlling for changes in the occupational structure of society) remained the same till 1940, at least in the province of Utrecht. Since then, several studies have been published about the development of marriage mobility in the Netherlands, also in comparison with other countries. More in general, it was shown that both for the country as a whole, as for urban and countryside areas, the association between social mobility and economic modernization could be researched and differentiated. Van Leeuwen, Maas, and Mandemakers (2005) showed that endogamy decreased in the second half of the 19th century, because children from the diminishing group of farmers had to move elsewhere to find marriage partners.

Bras and Kok (2005) researched marriage mobility to measure which marriages were homogeneous or heterogeneous in terms of social class background by comparing the social background of fathers and fathers-in-law. They used one of the first versions of the GENLIAS Zeeland dataset with marriages from 1796 till 1922. They did not find much influence of parental characteristics on having a heterogeneous marriage or not, except for the farmers who showed a strong inclination to marry in their own group (about 60 to 70%).

Kok and Mandemakers (2008) concentrated on the marriage market as a mechanism of partner selection and concluded that personal circumstances such as poverty, illiteracy and low social status decreased the chance of finding partners outside the local community. This effect was stronger in geographically more isolated communities. Concentrating on farmers, they found that it were wealthy farmers who were most able to place their children on a farm of their own, not through the choice of a successor to the existing farm, but by negotiating and bringing together the assets for a new farm household. This transmission required that both the father of the groom and bride were still alive (Kok, Mandemakers, & Damsma, 2010).

On the basis of a LINKS dataset of six provinces, Maas and van Leeuwen (2019), tested the modernization hypothesis again and confirmed that the occupational status of the father had become less important for partner selection in the second half of the 19th century. In cities, they found evidence that achieved characteristics became less important. This is a partial confirmation of the 'romantic love' hypothesis stating that both achievement and ascription became less important in partner selection.

### 3.2.3 STATUS ATTAINMENT

One of the first LINKS datasets consisted of the linked marriage certificates of five provinces of the Netherlands including about 0.4 million certificates. This dataset was widely used by sociologists to test the modernization theory which states that modernizing and industrializing societies become more open in terms of social mobility. Knigge (2015) used several status attainment models not only looking at the direct relationship between fathers and sons but including the association between brothers in his models as well. Covering the period 1827 to 1897 he found that the status correlation between father and son was relatively high (0.57) compared to 20th century societies. Remarkable was the high sibling correlation of 0.53 of which only 0.32 could be explained by the effects of the father alone. This implies strong family effects not directly passed on by the father. Parts of the family effect shown by siblings could be explained by direct effects of grandfathers and even great-grandfathers and these were relatively stable effects which endured over time and were independent of the chance of having direct or only indirect contact with the siblings (Knigge, 2016). The average effects decreased gradually after 1850 under the influence of some but not all modernization processes that changed Dutch society; after 1850 only the effects of urbanization and mass transportation showed significant results. Sibling cohesion proved to be weaker when age differences were relatively high, the status of the father shows more fluctuations and migratory moves interrupted the effect of birth ranking (Knigge, Maas, & van Leeuwen, 2014; Knigge, Maas, van Leeuwen, & Mandemakers, 2014). Sprok (2013)

found a negative effect of the number of siblings on one's social status, especially brothers. Bras, Kok, and Mandemakers (2010) also studied siblings' effects on status attainment through the perspective of resource dilution. They showed that the effects clearly depended on the social and temporal setting.

Zijdeman (2010) researched status attainment in several ways, using amongst others the LINKS dataset of the province of Zeeland covering the period 1811–1915. He not only included micro-data but also a large number of contextual data to control for changes in environment on the status attainment process. For the father-son relationship he found a more or less linearly decreasing association in social status; for the father-in-law relationship this was the opposite: as soon as the context became more industrialized and mass transportation developed, the association with the status of the father-in-law started to increase. However, as soon as the educational expansion became serious, the association with the parent-in-law began to decrease again (Zijdeman, 2009; Zijdeman & Maas, 2010).

To study the influence of religious background on life courses a special sample with an oversampling of Jews was made for The Hague. It was concluded that the 19th century social differences between Protestants, Jews and Roman Catholics were a consequence of a slowly changing class-based society and not of discrimination, at least the chances to change the social position proved to be the same for all religious groups (van Poppel, Liefbroer, & Schellekens, 2003). Within the context of a study into the occupational structure of Jews in Amsterdam between 1850 and 1940, Tammes (2012b) investigated the effect of different decisions regarding Jewishness on social status. He found that marrying a gentile or converting to Christianity had no effect on social status, however giving up one's religion was strongly associated with upward social mobility.

### 3.2.4 SOCIAL AND ECONOMIC CAREERS

Career mobility, the possibilities for persons to change their occupational position during their lives, could also be studied by using the HSN database. Maas and van Leeuwen (2009) researched to what extent occupational careers between 1865 and 1940 were to be considered as stable, upwardly or downwardly mobile both for males and females. They concluded that over the whole period mobility increased, both upward and downward. Where the average level of the occupational structure became higher as well, the male's total upward mobility equalled downward mobility. For females they found net upward mobility, although this could probably be explained by selection issues whereas females with lower estimated occupations disappeared from the registrations when they grew older. So, the influence of the social background did not change where it was expected that this influence would become less important over the years.

In her doctoral thesis, Schulz (2013) investigated the development of individual careers during the process of modernization and industrialization in the Netherlands. She tested if the role of the parents in the status attainment process diminished in favor of the role of education. She used the HSN dataset by employing multilevel growth models and found that career mobility decreased at older ages and that a father with a higher social status provided a much better start than other fathers, although this effect diminished over time. However, for females this was less clear and — unexpectedly — it turned out that being married had a positive effect on occupational status (Schulz & Maas, 2010). When looking at the development over time by comparing three cohorts active between 1865 and 1920, she found that intragenerational effects became increasingly more important for successful careers whereas the influence of social background diminished. Most mobility occurred at the start of the career, although for females this was less clear than for males maybe as a consequence of underreporting, especially by lower status groups (Schulz & Maas, 2012). Modernization indicated through, for instance, the level of educational opportunities and the number of steam engines proved to have no positive influence on career mobility, and the proximity of train stations even had a negative effect (Schulz, Maas, & van Leeuwen, 2015). More in general, females having a career proved to be more successful than males (Schulz, 2015).

The HSN was also used to zoom in on specific occupational groups. Boonstra (2011) used the HSN to study the intergenerational mobility of teachers, searching for real 'educational' families. He found indeed some 'hereditary transmission', but in general he concluded that it was a quite open profession during the 19th and early 20th century especially attracting newcomers from lower social backgrounds. Heerma van Voss and Vermeulen (2000) compared the life courses of active union members in the textile industry from the Twente region with a control group and concluded that the union members showed more upward social mobility. This could be explained by the large group of workers with a farming background in the control group. For them going back to a career in agriculture

was an opportunity that did not disappear when they started working in the textile factories. Finally, Bras (1998) studied females working in domestic services and concluded that this typical job, while it resulted in more geographical mobility, did not result in a higher social status except when associated with long-distance migration to the cities.

### 3.2.5 EDUCATION AND SOCIAL BACKGROUND

It was already known that literacy levels were relatively high for the Netherlands, compared with other European countries. In general, Protestant countries showed high levels and Roman Catholic countries rather low levels. In the Netherlands, the Roman Catholics who lived in Protestant-dominated areas caught up with the Protestants. Only the two southern provinces with a population of about 90% Roman Catholics, Noord-Brabant and Limburg, showed relatively low levels of literacy around 1800 (van der Woude, 1980).

Boonstra (1995) used the signatures in the first entered marriage certificates to make an overview of the development of literacy in the province of Utrecht. He concluded that at the start of the 19th century illiteracy had already receded to the lower social economic classes and to some remote areas in the province. After the middle of the century, illiteracy was not a factor of importance anymore. In a second study, Boonstra (2009a) could use the complete set of HSN birth certificates covering the whole of the Netherlands. Based on whether or not the father was able to sign, he studied the development of illiteracy from the birth year 1775 onwards. Illiteracy dropped from an average 30% to almost 0% for persons born in 1880. In this development the males took the lead. As said, in the early 19th century illiteracy levels were already very low in the north of the country except for the villages and towns that specialized in sea fishing. Boonstra also showed that the peat areas and the very conservative orthodox Protestant areas had rather low literacy levels as well.

Boonstra (2009b) extended his study of illiteracy and social and regional background into social mobility by combining the data on signatures and occupations from the certificates. He found that chances of both downward and upward intragenerational mobility were not influenced by being literate or not. However, on the intergenerational level it was important: being a son of a literate father gave a significantly higher chance of upward mobility and having an illiterate father gave a higher chance of downward mobility. Vandezande, Matthijs, and Kok (2011) tested the so-called resource dilution hypothesis on education by combining data on the number of children in a household and the capability to sign a marriage certificate. Having more brothers lowered the chance of being literate. The number of sisters had a different effect in the sense that having older sisters improved the chance of being literate for the younger sisters, but not for the younger brothers.

Zijdeman and Mandemakers (2008) researched the relation between social mobility and education. They compared a dataset with marriage certificates from pupils with a higher general secondary education background (HBS and Gymnasium) with a control group from the HSN database. They concluded that the direct importance of the parental background for the later social position of sons became less with a constant indirect influence by way of the educational achievement between 1880 and 1920. The influence of the educational achievement as such on the social position at marriage remained constant. The same dataset was used in an investigation of the impact of the HBS system on the labor market for white collar and professional occupations (Schalk, 2015).

### 3.2.6 GENDER, WOMEN AND WORK

During the 19th century the ideal of the 'male breadwinner' and the 'housewife' spread through the Netherlands. Van Poppel, van Dalen, and Walhout (2009) researched the way this diffusion of a cultural ideal took place on the basis of a GENLIAS dataset of marriage certificates. They found that the ideal diffused from the upper to the lower classes starting around 1830 in the urban areas and around 1850 in the rural areas of the Netherlands. Around 1830 on average about 30% of the marriage certificates from the different layers in the lower classes showed brides without an occupation at the moment of marriage, around 1910 about 60%, and within the group of farmers, 80%. For the upper classes this had always been the norm with a percentage that moved from 90% in the beginning of the 19th century to 97% around 1880 to start declining back to 90% around 1910.

Schulz, Maas, and van Leeuwen (2014) continued research on this question with the same dataset, concluding that not only the inclusion of females in the labor market dropped until about 1885 but

also the social prestige of the work they practiced. This trend was reversed by growing educational participation and dissemination of more egalitarian gender values after 1900, females changing from occupations such as housemaid to primary teacher, telephone operators, clerks, etc. On the basis of an oversampling of the HSN for the textile city of Enschede, Janssens (1998) researched the relationship between social background, occupation and religion in the female life course. Her main conclusion was that religion was the main factor in decisions regarding marriage and fertility and not the type of work in which women were engaged.

Van Leeuwen and Zijdemans (2010) used the Zeeland marriage certificates in combination with all kinds of data on the level of the municipalities. Their main question was how society became more open in terms of social mobility during the modernization of society. They showed that ascribed characteristics became less important while the effects of achieved characteristics were stable in the process of marrying a bride with a relatively high social status. Although the 'logic of industrialization' still proved the moving factor in this process, it turned out that on the municipal level characteristics of industrialization were less systematically related with changes in ascription and achievement than one might expect.

The LINKS marriage certificates dataset was also used to research why there was a decreasing female participation labor rate in the Netherlands in the course of the 19th century (Boter, 2017; van Poppel, van Dalen, & Walhout, 2009). As both brides and grooms were asked to state their profession at the moment of marriage, it was possible to research this phenomenon on the micro-level. Cultural reasons like the advance of the male breadwinners ideal and the increased income for large parts of society made it for women less necessary to work. Where van Poppel, van Dalen, and Walhout used only the provinces and period as independent variables, Boter researched the developments at the municipal level. She showed that also the changing economic structure was causing diminishing female labor participation, whereas in some areas the reverse trend was visible (e.g. textile industry).

For her dissertation, Bras (2002, 2004) concentrated on the career mobility of maids from Zeeland and found that for half of them the career did not last longer than five years and that there were no serious mobility routes in terms of the social background of the employers. The social background of the maid was also positively related with the social background of the family in which she got a position. In general maids from middle class background migrated more to cities which gave them higher chances of upward social mobility and gaining better living conditions.

### 3.3 DEMOGRAPHIC TOPICS

#### 3.3.1 MORTALITY ACROSS THE LIFE COURSE

Since the middle of the 19th century scholars have debated the effect of industrialization on living standards, for which (infant) mortality quickly came to be seen as a good indicator. 'Pessimists' predicted that increasing inequality during industrialization — combined with the receding effects of epidemics — would lead to increased social class differentials in mortality. 'Optimists' pointed at higher real wages, which would soon be translated into better housing and food, and in a lower social gap in mortality rates. Yet others predicted that social class differences would remain constant over time, as the rich would always be able to translate their resources into better health (Bengtsson & van Poppel, 2011). Already from the start, HSN was used to add to this debate. The first papers dealt with infant and child mortality (e.g. van Poppel, 1995; van Poppel & Mandemakers, 1997; van Poppel & Mandemakers, 2002; Woods, Løkke, & van Poppel, 2006). The fact that the data covered large regions, with distinctly different health environments and traditions, made it possible to study class differentials in infant and child mortality across regions. Van Poppel, Jonker, and Mandemakers (2005) discovered clear social gradients in both infant and child mortality. However, the region in which a child was born was more important than the class position of its father. Contrary to what was expected, they found that in a 'healthy region' (higher quality of surface water), in combination with breastfeeding practices, social class differences were limited compared to more unhealthy regions. A recent study using the LINKS data confirmed that limited breastfeeding in Zeeland, associated with female field work, led to an outspoken pattern of high infant mortality in the first summer after birth (van Poppel, Ekamper, & Mandemakers, 2018). High temperatures would increase mortality risks even more (Ekamper, van Poppel, van Duin, & Garssen, 2009; Ekamper, van Poppel, van Duin, & Mandemakers, 2010).

Infant mortality is also the subject of a long-term debate on the role of the Roman Catholic clergy, who, supposedly, in the late 19th century began to discourage mothers to breastfeed their babies thus causing higher infant mortality in Catholic regions (on religious differentials, see also van Poppel, Schellekens, and Liefbroer (2002)). Supposedly, this prudishness was part of their 'moral campaign': in their fierce competition with Protestants, Roman Catholics aimed to demonstrate they had the highest moral standards. Although this particular role of the clergy has recently been put into question, it is still possible that Catholic mothers were less inclined to breastfeed their infants. Inspecting birth intervals by religion, Janssens and Pelzer (2014) concluded that in several places (in the period 1880–1920) Catholic mothers were less likely to breastfeed than mothers from other denominations. However, this did not translate into Catholic excess mortality. In their view, region was the most important factor in explaining infant mortality. But in her recent doctoral dissertation, Walhout (2019) concluded that religion, after all, superseded region in explaining the higher infant mortality risks of Catholic babies.

Illegitimacy has always been related to high risks of dying in infancy. The question remained whether these high risks could be fully explained by the social conditions of the single mothers — poverty, exclusion from poor relief, social isolation and inability to combine work with breastfeeding — or whether infanticide was also involved. By comparing with poor mothers, van Poppel, Kok, and Kruse (1997) concluded that infanticide, or at least, willful neglect played an important role in high mortality of illegitimates.

Social class also appears to be the most important factor in explaining excess mortality of teenage girls, a phenomenon that occurred in most European countries before 1940. In the Netherlands, the problem seems to have been limited to daughters of unskilled workers. Van Poppel, Schellekens, and Walhout (2009) explain the relatively favorable situation for Dutch girls from "the dominant position of small family farms, the preponderance of mixed or dairy farming, the well-integrated position of women in market production and, more generally, the higher degree of equality between men and women" (p. 37). In contrast, the wives of farmers and skilled workers ran higher risks of maternal mortality than women from other social groups, probably because women in these groups tended to perform heavy labor in advanced stages of their pregnancy (Ory & van Poppel, 2013).

As to mortality at older ages, social class differences turned out to be even less conspicuous. Although somewhat paradoxical, van Poppel, Jennissen, and Mandemakers (2009) found relatively strong differences for the provinces of Zeeland and Limburg. Van Poppel and van Gaalen (2008a) concluded that there was no sign of living standards of workers deteriorating during industrialization and urbanization which they interpret, rather reluctantly, in line of the 'optimist school': the lack of class differences relate to relatively high real wages and a quite generous poor relief in The Netherlands. Their results were corroborated by Schenk and van Poppel (2011), who state that Dutch economic growth benefited the lower classes, who "took advantage of the new possibilities created by increased medical knowledge, improved sanitary standards, which in their turn were not independent of the increased economic growth" (p. 415).

### 3.3.2 HOUSEHOLD COMPOSITION AND ITS DEMOGRAPHIC IMPLICATIONS

One of the oldest topics in historical demography is household composition, for which the Cambridge Group for the History of Population and Social Structure created elaborate schemes to make distinctions between types of household extensions (Hammel & Laslett, 1974). One of the most important findings in this line of research was that the nuclear family was not the outcome of processes of industrialization and urbanization, but that it had antedated them by at least several centuries. The Netherlands clearly fits into the English pattern of nuclear families, but the HSN made it possible to study regional differences and the role of agriculture in household extension in greater detail. HSN data allowed to map two regions of extended families. One along the eastern border with Germany where a stem family system persisted, even in the face of legally prescribed equal division of inheritance. And another one in the south, which was related, ironically, to strict partibility coupled with struggling small farms. In this region, adult (unmarried) children tended to stick together in order to hold on to their share (Kok & Mandemakers, 2009). The HSN also makes it possible to look at extended households from the perspective of persons potentially in need of care, such as orphans, unmarried persons, widows, widowers and the elderly. Research showed that many of those found recourse, at least temporarily, in the homes of family. Especially in the northern and western regions, more commercialized and urban areas of the Netherlands, household extension was often related to a crisis situation. E.g. a couple took

in an ageing parent who could no longer live independently. This research put the notion of 'nuclear hardship' in perspective (Kok & Mandemakers, 2010, 2012).

Another approach to households in the HSN is to take a long-term perspective and to compare the (prospective) data from the HSN with (retrospective) panel surveys. In this way, van Poppel, Schenk, and van Gaalen (2013) created a continuous birth cohort from 1850 until 1985. They studied the living arrangements of children, and were able to dispel the notion that nowadays more children are living in broken or reconstituted families than ever. They discovered that, due to the decline in adult mortality, the decrease in out-of-wedlock fertility, and rather low divorce frequencies, the cohorts born between 1900 and 1965 were spending more time with near kin than "at any time in history" (p. 255). Also, the traditional, complete nuclear family in which children live with both their biological parents is nowadays still stronger than it was in the middle of the 19th century (see also van Poppel, Tammes, and Schenk, 2010; on urban-rural differences in the living arrangements of children, see also van Gaalen, 2007; van Gaalen and van Poppel, 2009). Lin (2011) studied female headship using the HSN data for Rotterdam, and contrasted it to patterns in Taipei, Taiwan. Incidence, timing and duration of female headship differed strongly, which Lin sees as a (new) indicator of differences between European and East-Asian family formation systems.

In his recent dissertation, Rosenbaum-Feldbrügge (2019) used the HSN to study the effects of parental loss on a number of outcomes: the remarriage or migration of the widow or widower (also Rosenbaum-Feldbrügge, 2018), and the impact on the timing and incidence of leaving home and marriage of the children. Furthermore, he looked at the long-term effects of bereavement on the occupational status of (half-)orphans in early adulthood. The findings support to some extent the 'niche' hypothesis that the marriage of sons of farmers and skilled workers was advanced by parental death (Rosenbaum-Feldbrügge & Debiasi, 2019). The study shows that the death of the mother had more devastating and lasting effects than the death of the father, e.g. in terms of status attainment. The results point at the more pivotal role mothers played in care and upbringing — combined with the generous Dutch poor relief which may have made widows less vulnerable than might be expected.

### 3.3.3 NUPTIALITY

The study of marriage frequency and timing has always been core business of historical demography. How was marriage related to the requirement of an independent household? Why did property-less people seem to abandon traditional restraint in the second half of the 19th century and began to marry more often and younger? To be sure, the data collection of marriage certificates in the HSN was soon superseded by the massive collection of marriage data available through the LINKS project which allowed not only the study of marriage timing, but also of intergenerational transmission of the age at marriage (van Poppel, Monden, & Mandemakers, 2008), the impact of birth order (Suanet & Bras, 2014), the development of the marriage market in a geographic sense (Ekamper, van Poppel, & Mandemakers, 2011) and the seasonality of marriage as a marker of secularization (Engelen, 2017).

One of the first possibilities with the LINKS dataset of marriage certificates, was the study of kin marriages (Bras, van Poppel, & Mandemakers, 2009). Kin marriages were rather common among farmers and the higher social strata, clearly as a way to consolidate and merge properties. The increase of cousin marriage among orthodox Protestants in the Bible Belt area of the Netherlands was interpreted as a form of identity preservation. Overall, kin marriages took place more often in the relatively isolated, inland provinces of the Netherlands. Finally, sibling set exchange marriages were a consequence of the enlarged supply of same-generation kin as a result of the demographic transition.

What could the more limited collection of marriage certificates in the HSN add to the impressive LINKS findings? Of course, the HSN datasets offers much more variables than available in the LINKS database. To begin with, Kalmijn (1995) used the first HSN release to link the decreasing marriage age with social class, urban areas and migration. Van der Velden (2012) used HSN data to compare the age of marriage of seafarers with the inland workforce and concluded that marriage ages did not differ widely while he had expected that seafarers would marry later than the inland males. Other groups of whom marriage behavior was studied include farmers (van Poppel, Ekamper, & van Solinge, 2007) and workers in Rotterdam (Kok, 2006a). The HSN was also employed to relate timing and incidence of marriage to the RPs family situation both early in the life course (Engelen & Kok, 2003) and at age 18 (Kok, 2014b). Thus, these researchers assessed the influence of religion, living in a city, region, social economic status, and the family composition on the likelihood to marry early, at a normal age, late

or not at all. This procedure also allowed to answer the question how late marriage and permanent celibacy were related; did the unmarried persons have the same 'profile' as those who married late? It was found, that indeed several groups of singles (e.g. youngest sisters) simply had waited too long in seeking a marriage partner. But others had a different profile, especially women from the upper classes, who could not combine education and fitting work with marriage. Finally, in the eastern part of the country late marriage was normal, but celibacy was not. Marriage was sure to come, but it depended on the timing of succession (Engelen & Kok, 2003, p. 27; on the life courses of urban singles, see Kok & Mandemakers, 2016). A more sophisticated way of studying life course 'pathways' into family formation, and the shift in these pathways of time was performed by Bras, Liefbroer, and Elzinga (2010). They used sequence analysis to uncover life course trajectories into adulthood, and could demonstrate the gradual emergence of the typical mid-20th century 'standard' life course of early family formation. Forerunners in this process appear to have been "laboring class youths, farmers' daughters, youngsters of mixed religious background, and the urban-born" (p. 1030).

As mentioned earlier, HSN procedures have been applied in specific projects. An example is the reconstruction of two complete generations (including migrants) of the people who married in the North-Holland rural municipality of Akersloot (first generation), which was selected primarily because it had already started a population register in 1830. The sample has been used to study, among others, the effects of the differences in sibling sets on marriage chances. Bras and Kok (2016) found, among others, that resource dilution in this community was gender-specific: "the more sisters in the household, the more the marriage of boys and girls is postponed or forfeited. Clearly, girls meant a drain on the resources at home" (p. 202). Another research using this dataset studied the intergenerational transmission of the age at marriage in relation to social control. The hypothesis is that such transmission effects are strong in groups with weak social control mechanism. This hypothesis was supported by the finding that mother to daughter transmission was weaker among farmers and among Roman Catholics than among working and middle classes and Liberal Protestants (Kok & Van Bavel, 2009).

The Akersloot database has also been used to study partner choice and social reproduction among farmers (Kok, Mandemakers, & Damsma, 2010). Partner choice was also the subject of research based on an oversample of Amsterdam Jews (religious intermarriage, Tammes, 2010) and a sample of Germans living in Utrecht (intermarriage with other ethnicities, Schrover, 2001, 2003, 2004, 2005). A large special project entailed the reconstruction of the life courses of Germans, Italians and persons from the Dutch provinces Zeeland and Noord-Brabant migrating to Rotterdam. Two migration cohorts were created which were followed up to the third generation. This research into assimilation processes obviously also dealt with the likelihood of marrying natives, people from the home region, or other migrants (Lucassen, 2003, 2005; Chotkowski, 2006).

Partner choice studies using the HSN have dealt with the various preferences for endogamy (by locality, religion, and class) and how they interact (Kok & Mandemakers, 2008; using GenLias, Ekamper, van Poppel, van Duin, & Mandemakers, 2010; Maas & Zijdeman, 2010; Zijdeman & Maas, 2010). The question whether family pressure leads to more social endogamy was studied by van Leeuwen, Maas, and Mandemakers (2005) who looked at the presence of parents and other family as witnesses at the wedding. They conclude that in general, farmers excepted, there was little family pressure leading to endogamy. More recent studies take up the theme of migrant integration through marriage again. The city of Rotterdam is compared to Stockholm and Antwerp to enable a more detailed study of how characteristics of cities enable migrant integration (Puschmann, Van den Driessche, Grönberg, Van de Putte, & Matthijs, 2015). The authors conclude that migrants faced many difficulties and social exclusion, resulting in low chances of marrying natives. The situation in Rotterdam and Antwerp appears to have been better than in more industrialized Stockholm, which suggests that port cities do offer some routes to escape exclusion.

Marriage can of course be triggered by a pregnancy (see Kok, 2011 on forced marriages in HSN). Bridal pregnancies in relation to regional courtship norms are the subject of a recent paper by Kok, Bras, and Rotering (2016). Local customs, however, were not strongly related to levels of bridal pregnancy. The phenomenon was strongly concentrated in proletarian as well as in Protestant groups. There is evidence of parental tolerance for the sexual urges of (endogamous) young couples who posed no threat to the property transmission. But there are also indications that youths deliberately advanced a marriage (using a pregnancy as leverage) to gain independence.

Recently, Jennings and Gray (2017) related marriage data from the HSN to climate fluctuations culled from the archives of the Dutch meteorological institute. Contrary to Malthusian expectations, people did not marry because they felt secure of their prospects. On the contrary: "Adverse environmental and food security conditions consistently increase marriage in this historical context" (p. 257). The authors explain this outcome from intra household competition over scarce resources, a situation in which children might see leaving home and marriage as a way to improve their situation. Possibly rural children migrated to cities where it was easier to find a partner and marry.

Finally, HSN has been used to study remarriage (van Leeuwen & Maas, 2007b) and divorce. Kalmijn (2008) noted a clear social gradient in the likelihood of a marriage ending in divorce. Forerunners of the recent divorce explosion appear to have been the 'cultural elite'. Higher professionals, but not managers, were more likely to divorce than others. He notes that the lower classes might have chosen other options, such as simply leaving, to end an unhappy marriage.

### 3.3.4 FERTILITY

As elsewhere, entire generations of Dutch historical demographers have been puzzled by the fertility decline, and we can safely say that HSN has enlivened but not concluded the debate. Dutch demographers faced the challenge of having to explain two conspicuous features of demographic change. First, the onset of the transition had a marked regional pattern, with the north-western parts of the country being the first to show change and the south-eastern parts the last. Second, within Western Europe, the Netherlands as a whole was a latecomer, characterized by relatively high fertility levels until the 1960s. Initially, the explanation followed a classic modernization discourse, in which fertility was held in check by late marriage, as household formation was contingent on inheriting a farm or workshop. The spread of wage labor supposedly led to an 'intermediate' period of earlier marriages and high fertility, until people adopted modern birth control techniques to limit family size (for an overview of the debate, see Engelen (2009)). The willingness to do so presupposed an attitude of self-determination, which was part of modern culture, which supposedly first emerged in the North and West, and gradually spread across the country. Critics of this view claimed that the regional pattern in Dutch fertility decline could be explained more convincingly by the predominance of Roman Catholics in the South, in combination with the Catholic opposition to birth control. Throughout the 1960s to 1980s, scholars have tried to solve the debate by weighing the relative contributions to fertility decline of social and economic 'modernization' on the one hand and religion on the other.

Internationally, the debate has been between those who see fertility decline as an 'innovation' of new ideas and techniques, which means research had to focus on the venues and mechanisms of diffusion, such as social learning. On the other side, scholars still see merits in the older conception of fertility decline as a response to decline in mortality.

Ethnographers have for a long time documented the contraceptive methods employed in non-Western societies. Recently, historical demographers have demonstrated that conscious birth spacing, for instance due to expected increases in food prices, was practiced by pre-transition European populations (e.g. Van Bavel, 2001; Dribe & Scalone, 2010). Birth spacing intensified during the first stage of the fertility decline, for example through protracted breastfeeding. Even in the 20th century, traditional contraceptive methods such as abstinence and withdrawal remained the most important techniques for married couples. The HSN has contributed to this field by providing family reconstitutions which were not, as usual, limited to specific regions or social groups. Moreover, the database makes it possible to weigh the relative impact of socio-economic and cultural factors (especially religion) at the level of individual couples. And not only parity-specific stopping behavior was studied, but also spacing and childlessness. Probably the most important contribution of register data to the study of fertility is the possibility to study the conditions triggering agency or the willingness and ability to limit one's fertility.

Spacing studies using the HSN have revealed that (working class) couples burdened with many young children delayed the arrival of a next child (Van Bavel & Kok, 2004, 2010c). However, a deliberate response to rising food prices was not found (Van Bavel & Kok, 2005b). Time and again, religion proved a prime factor in determining whether people exercised agency to control births or not (Schellekens & van Poppel, 2006). Especially Roman Catholics stood out with short birth intervals, which may be related to a lower incidence of breastfeeding, the 'conjugal duty' (spouses were not supposed to refuse to have sex) or active pronatalism (Van Bavel & Kok, 2005b). Recently, van Poppel, Reher, Sanz-Gimeno, Sanchez Dominguez, and Beekink (2012) demonstrated that couples actively responded to

deaths of their children, in other words spacing and stopping was contingent on childhood survival (also [Reher, Sandström, Sanz-Gimeno, & van Poppel, 2017](#); [Shepherd, Kok, & Hsieh, 2006](#)). Thus, they step in on the 'adaption' side of the fertility debate. They observed that "childhood survival had clear effects on reproduction, the chances of having another child, and the length of the intervals between births, which indicates that this variable was crucial for fertility decisions" (p. 300). The pattern was especially strong after 1900. This form of adaptive agency was not found in all groups. Skilled workers and Liberal Protestants show the reaction much stronger than farmers, Roman Catholics and Orthodox Protestants. A similar conclusion was reached by [Schellekens and van Poppel \(2012\)](#) who attribute the decline in fertility in the Netherlands before 1940 to the decline in child mortality, rising real wages and unemployment during the Depression. These see no room for innovation/diffusion models: "ideational change that is independent of social and economic change was not a major determinant of the decline" (p. 982).

The Interwar period has also been subject of research into the remarkable rise in the number of couples who had no, or only one, child. With hindsight, low fertility in this period has been interpreted as the effects of unemployment and looming war, but contemporary commentators were blaming the remarkable increase of childless or one-child families on changing life styles. Supposedly, many couples were hesitant to give up newly-won luxuries and holidays, and supposedly decided to have no or few children. To check whether childlessness was a reaction to adverse circumstances or a reflection of a 'modern lifestyle', [Van Bavel and Kok \(2010a\)](#) combined the reproductive histories of almost 3,000 HSN couples with information on their municipalities (e.g. level of unemployment, number of stores). Although, indeed, married men without occupation often remained childless, they also found indications that more or less deliberate childlessness was involved in order to defend a career or a modern lifestyle. Controlling for age at marriage, white-collar workers had relatively high hazards of childlessness, although the difference with other groups of workers gets small and statistically insignificant after controlling for regional and municipal characteristics. As to the municipal effects, people in localities with many stores tended to have no children, suggestive of a 'luxury effect'. The same applied to people in urban environments. Furthermore, couples who had a mixed marriage or who were both without a religion had relatively high odds of childlessness (also [Van Bavel & Kok, 2010b](#) and [Van Bavel, Kok, & Engelen, 2008](#)). The 20th century also witnessed the emergence of the ideal of a mixed offspring set, consisting of a boy and a girl. [Kok \(2018\)](#) showed that this emerged first among Liberal Protestants.

The HSN has also been used to demonstrate that there is still room for a 'blended' approach to fertility decline in which diffusion of new ideas and attitudes plays a role alongside structural factors. [Bras \(2014a\)](#) studies social networks of HSN RPs by looking at the witnesses at the marriage ceremony. Age, occupation and relationship to bride or groom of these (four) witnesses were recorded. [Bras](#) discovered that when witnesses consisted of siblings, age-peers and/or women, the couples in question were more likely to practice birth control. She sees these reference groups as "salient role models for social learning" (p. 178). These 'controlling' couples stand in contrast to couples who had more 'vertical' ties, that is witnesses from their parent's generation — in particular unskilled workers and farm workers. Using an oversample of four industrial cities, [Janssens \(2009; 2014\)](#) studied the impact of female working experience before marriage on their reproduction. She assumed that women coming into contact during their work with upper classes (such as domestic servants), or women working alongside other women (factory girls) would have had more opportunity for social learning than, for instance, women working at home. She showed that (former) servants were actually late in adopting birth control. Interestingly, a high education before marriage did not have the expected effect of lower fertility (through an increased bargaining position with the husband) either.

Most evidence from the HSN points into the direction of fertility decline being an adaptation to changing circumstances, in particular declining child mortality. But there are also indications that some ideational change was going on, diffused through networks of peers. However, the old Dutch debate on the striking regional variation in fertility decline in the Netherlands is still far from solved.

Apart from explaining socio-cultural variation in fertility and its decline, innovative research using longitudinal micro-data has come from molecular epidemiologists and researchers testing hypotheses from evolutionary biology. A large oversample of Rotterdam women, who were traced from birth to their own families has been employed in epidemiological studies to discover negative effects from the season of birth ([Smits, Jongbloet, & Zielhuis, 2001](#); [Smits, Zielhuis, Jongbloet, & Straatman, 1998](#)); the age of RPs mother ([Smits, Zielhuis, Jongbloet, & van Poppel, 2002](#)) and the interval after which

the RP was born (Smits, Jongbloet, & Zielhuis, 2000) on her own fecundity (also Smits, 1998). With respect to the preceding birth interval, it was discovered that RPs born after very short birth intervals (less than 1 year) showed higher likelihood of childlessness and stillbirth in their offspring, compared to those born after intermediate intervals (21–32 months). The age of the mother was important as well: RPs whose mother was over forty had higher risks of stillbirths, childlessness and multiple births than RPs born from a mother of intermediate age (24–30). The explanation lies in ovarian maldevelopment, which already occurs in fetal stage. Recently, the LINKS data were used together with other databases on 'natural fertility' populations to chart the distribution of the final age at childbirth (Eijkemans, van Poppel, Habbema, Smith, Leridon, & te Velde, 2014).

In recent studies, the relevance of evolved human traits for understanding (historical) demographic behavior has been tested. One example is the notion that the human race stands out compared with other animals by the role kin groups play in rearing children. It has been suggested that through evolution, women have developed a relatively early menopause in order to help their daughters with raising children. The notion of co-operative breeding or the 'grandmother hypothesis' has been tested using data from several villages, studies in developing countries and historical family reconstitutions (Sear & Coall, 2011). Overall, maternal grandmothers appear the most reliable 'helpers in the nest'. This hypothesis was tested using household information contained in the HSN by Rotering and Bras (2015). They studied the effect of living-in kin on birth intervals, and expected that 'helpers in the nest' would lead to shorter intervals. Some kin effects were found, but only in the low parities. Also, a 'grandmother effect' was not found. Living-in widowed grandfathers even had the opposite effect of delaying the arrival of a next child, probably because he was using up too much of the couple's resources. As to siblings it was not, as could be expected, sisters who would stimulate fertility, but brothers — apparently because they brought in additional income.

## 4 NEW RESEARCH FIELDS

In the wake of poststructuralism and responding to renewed interest in individual agency, the early 1990s saw the emergence of new topics, such as individual life plans, individual and family strategies of betterment, etc. Researchers began to use the HSN to study how people responded (e.g. by the timing and reason for leaving home, and the timing of marriage) to parental needs in relation to their own preferences. Apart from parental social class, family size and sibling order were seen as key indicators to study to what extent demographic choices were determined by parental resources and family constraints. HSN researchers could also join emerging new fields of research, such as the role of siblings and the wider kin network in influences on life courses, the topic of 'death clustering' at the level of families, and the study on how early life conditions affected health in later life.

The HSN's rather unique research design — following individuals through the entire country, and even beyond — also made it the perfect database for in-depth studies of migration. Until the early 1990s, research into migration was always hampered by skewed comparisons (leavers contrasted to stayers, in-migrants contrasted to natives), and the fact that generally only one move per individual was observed, as research was always limited to one locality. Not surprisingly, the HSN stimulated diverse studies of migratory behavior.

### 4.1 MIGRATION

Being a servant as part of the life cycle, has always been an important element of the North-Western European family system, and as such, a key factor in historical demographic interpretations of late marriage (Hajnal, 1982). However, the 'life cycle servants' themselves have hardly been subjects of research: from what families did they come, at what age and why did they leave home, how long did they stay away, did they remit their earnings to their family, did their experiences away from home — e.g. living in urban, upper-class households — affect their choices later in life? Using the data on the first completed province (Utrecht), Kok (1997) charted the family background of leaving home, and traced the domestic servants and farm workers in their subsequent jobs (also Bras & Kok, 2003). Recently, the various destinations (ranging from an interlocal move to emigration) of rural youths leaving home were studied using a competing risk event history analysis (Mönkediek, Kok, & Mandemakers, 2015; Mandemakers, Mönkediek, & Kok, 2016).

Bras focused on girls from the province of Zeeland, and included sisters in her research project. Her aim was to find out whether and how life chances of women were affected by (possible) service, the location and employers, and the migration trajectory, controlling for parental environment. One of the hypotheses was that domestic service was a 'bridging occupation', allowing girls from rural lower-class families to achieve upward mobility because of their increased social and cultural capital gained in cities. However, service mobility in Zeeland remain predominantly local, with only a minority of the girls actually working in cities. Those who did were indeed able to marry a higher status husband, but local city girls fared even better (Bras, 1998; also Bras, 2002). Girls moving to a city were already a selected group, being able to draw on parental resources. The girls came from families of teachers, supervisors, lower civil servants, and skilled laborers. These families could afford losing direct income by having a family member working in the urban sector and could possibly spread their risks in times of economic hardship. They may also have seen a position in urban domestic service as an opportunity for education and upward mobility for their daughters. Siblings already living outside Zeeland enabled a move out of the province (Bras, 2003).

Sibling effects on the likelihood and direction of migration was also the topic of a comparative study contrasting Zeeland to the Pays d'Herve (Eastern Belgium). Bras and Neven (2007) demonstrated that the type of labor market, which differed strongly between these regions, affected whether women would be stimulated by their brothers or sisters to migrate. In Zeeland, girls were affected primarily by their sisters, because of the gender division in the labor market: girls were sent into domestic service, boys to work in agriculture. The marriage of a sister implied that another girl was sent into service, to make up for the loss of income. In Eastern Belgium, service was not common. Here, migration of females was influenced primarily by the size of the sibship, by the presence of younger siblings at age 12, and by the births and deaths of new siblings, in other words the parents aimed to regulate the size of the household. The effects of siblings were less sex-specific in the Belgian area. Since the women migrated into urban employment opportunities in Walloon industry, the contacts and resources of both their sisters and their brothers were important (Bras & Neven, 2007). Taking another approach to siblings' migrations, Kok and Bras (2008) used the Akersloot oversample (see above) to study how families dispersed. By looking at the residences of parents and their adult children in different stages they charted dynamic 'family territories'. Marriage was the prime factor in explaining dispersal of siblings.

The study of internal migration also included family migration, and the HSN register data allowed scholars to add previous migration experiences and the (dynamic) composition of the household to their models (Kok, 2004; Kok, Mandemakers, & Mönkediek, 2014). Detailed migration information spanning the life course is hard to come by, and the findings from HSN were also used to 'calibrate' migration studies based on genealogical information. As genealogies may be biased by the social class of the descendants reconstructing their family trees, and by the fact that they depend on vital events (marriage and child births), a check with register data can be useful (Kok, Adams, Ericsson, & Moch, 2002; Kok, Lucassen, Kasakoff, & Schwartz, 2002). In particular, a comparison was made with American genealogies, by simulating the Dutch data in such a way that it became clear how many moves were missed by reliance on vital events, and in what stage of the life course (Adams, Kasakoff, & Kok, 2002). Recently, Jennings and Gray (2014) related migration to weather conditions (time series of precipitation and temperature), thus relating to current concerns of the impact of climate change on migration. They showed that only internal mobility of specific social groups could be labelled 'distress migration', whereas emigration was actually limited by adverse conditions. Over time, people were less 'trapped' by bad weather conditions and moved more freely (Jennings & Gray, 2014). Finally, HSN also allows the study of residential mobility, a topic much neglected in historical demography. Kok, Mandemakers, and Wals (2005) analyzed the address changes of Amsterdam dockworkers, people on poor relief, and the HSN sample of Amsterdam. They interpreted residential mobility as a 'coping strategy'; when cheap dwellings were available, poor people moved frequently to save on the rent, which was only an option when the move was over a very short distance and did not entail a break with their social network on which the poor relied for survival. Van der Harst (2006) traced the moves and residential clustering of Zeeland and Brabant migrants in the city of Rotterdam.

By definition, foreigners are absent in the HSN sample, since it is based on the birth records. Life courses, however, are ideally suited to study integration and assimilation through, for example, residential clustering, chances of upward mobility and intermarriage. Following HSN procedures, cohorts of migrants have been followed through the registers in Rotterdam and Utrecht. The focus in these projects was on more permanent migration, whereas many immigrants soon moved on or returned,

probably because the city offered them no prospects. Perhaps the migration literature has paid too much attention to this 'floating proletariat' of migrants who left quickly (Thernstrom, 1973). Germans who stayed in Rotterdam actually fared very well. They were positively selected to begin with, but they managed to reach higher positions than the native born (Lucassen, 2004; also Delger, 2003; Schrover, 2002; van de Laar, Lucassen, & Mandemakers, 2006). Many Germans came to the Netherlands to occupy specific occupational niches (e.g. file-makers or stucco-workers), and their migrations are clear examples of chain migration, linking a few places in the German hinterland to Dutch cities. The study of German communities offered interesting insights in how these niches developed and how they remained dependent on incoming migrants from the home community, preferably both men and women (Schrover, 2000a, 2000b, 2001, 2003). However, the study of German immigrants has also revealed that many of them were not connected to other Germans, and apparently managed perfectly well to find housing and work on their own. Lesger, Lucassen, and Schrover (2002) conclude that ethnic 'chain migration' has received too much attention in the migration literature — in which American scholars predominated — probably because 'community studies' in the US were more attractive than studies of loosely attached migrants.

The study of (r)emigration on the basis of HSN is still limited. Following migrants into Germany is hardly possible, because of the lack of register data. Migrants crossing the border into Belgium have been followed, but they have not been studied separately. An important exception is a study into the careers of Dutch (HSN) men who joined the colonial army in the Dutch East Indies. The commonly held opinion is that such colonial armies consisted of the dregs of the society. It turned out that many of these migrants had an urban background, in contrast to the migration of the Dutch to the United States, which had a strong rural basis. However, the milieu of the unskilled urban proletariat was under-represented among the migrants to the Dutch East Indies. Apparently, engaging for colonial military service, even as an ordinary soldier, was considered to be a serious option for sons from the lower middle classes and artisans of the Dutch cities. The higher classes were actually over-represented among those who left for the Dutch East Indies (Bosma & Mandemakers, 2008; Bosma, 2009, 2010). Recently, an attempt has been made to trace HSN emigrants in the US censuses (Paiva & Anguita, 2017).

## 4.2 DEATH CLUSTERING

Increasingly, scholars have become aware that early-life mortality was often clustered in families. In other words, a relatively large proportion of all deaths occurred in a relatively small number of families. To properly understand the trends and local variation in infant mortality we need to know more about this phenomenon. Factors that may account for the clustering include the mother (her age, fecundity, inherited genetic disorders, disease profile, level of education, training in hygiene and breastfeeding habits), the household (income and composition) and the community (ecology and disease environment) (Edvinsson & Janssens, 2012). For the Netherlands, a study based on the LINKS dataset has shown that infant death clustering was unevenly spread, being very strong in the province of Zeeland. The study revealed that part of the explanation lies in the competition between young children for care, attention and resources within households. The mother's health was also important, which was indicated by a strong association between stillbirths and infant mortality; if a mother has poor health, she has more births with low gestational ages and low birth weights (van Poppel, Bijwaard, Ekamper, & Mandemakers, 2012). Using registers data, Janssens and Pelzer (2011, 2012) were able to attribute infant deaths in Enschede and Tilburg completely to the 'usual suspects' such as twin births, the preceding birth interval and the age of the mother. Therefore, there appears to be limited explanatory space for bad childrearing practices, at least in these two cities.

The recent dissertations of van Dijk (2019), Mourits (2019) and van den Berg (2020) all dealt with various aspects of death clustering. Van Dijk used extended family reconstitutions of Zeeland (based on LINKS) to trace the importance as well as the intergenerational transmission of familial death clustering (also van Dijk, Janssens, & Smith, 2018; van Dijk & Mandemakers, 2018). She found that children growing up in high-mortality families suffered lasting consequences in terms of lower survival chances of themselves and of their children. The mother's experience of sibling mortality turned out to be more important than the father's. Mourits (2019) focused on mortality after age 50 and individual chances to belong to the top 10% survivors of their birth cohort. He demonstrated that in longevity as well there was a clear element of familial clustering. This implies that 'average' life tables can be quite meaningless, as the distribution of survival chances are strongly determined by family effects.

In understanding longevity clustering, Mourits extensively explored social (resource transmission) and geographical components (see also [Mourits, 2017](#)). Finally, van den Berg (2019) used LINKS and HSN to develop new concepts and methods to study the familial component in longevity (also [van den Berg et al., 2019a, 2019b](#)). For his research, a case and control group design based on HSN was followed which implied tracing all descendants of exceptionally old RPs as well as a control group from late 19th century to the present day. This unique dataset contains five generations connecting living persons to their deceased ancestors. Van den Berg unequivocally showed that longevity is only transmitted as a quantitative genetic trait if persons belong to the oldest 10% survivors of their birth cohort when at least 30% of their ancestors also belonged to the oldest 10% survivors ([van den Berg et al., 2019a, 2019b](#)). A comparative study, covering five databases, including the LINKS database for the province of Zeeland, showed a strong relationship between levels of infant mortality of grandmothers and their daughters ([Quaranta & Sommereth, 2018](#)).

### 4.3 EARLY AND MID-LIFE EFFECTS ON HEALTH AND MORTALITY

Since the early 1990s, the notion that later life diseases already have their origins in the fetal stage (e.g. [Barker, 1992](#)) has stimulated a lot of research. A number of important contributions to this field were based on HSN register data. Early life conditions were studied for birth cohorts, which were exposed to economic crises or epidemic diseases. For the survivors, having experienced these adverse conditions could still imply a deterioration of health in later life. Van den Berg, Lindeboom, and Lopez (2006) looked at the effects of the business cycle in the years HSN RPs were born. They discovered that an individual born during a recession lived a few years less than an individual born in a boom. The outcomes suggest that the very first year of life is already crucial. As for the causal pathway, they suggest that the cyclical turbulence creates stress among parents, which in turn creates high mortality later in life among their infants. In a follow-up study, Yeung, van den Berg, Lindeboom, and Portrait (2014) merged the HSN life course data (cohorts 1880–1918) with cause-of-death information held at the Statistics Netherlands (CBS). This showed, surprisingly, that adverse condition in early life led to higher mortality due to cancer, as well to chronic respiratory diseases (females). Overall, the effects of economic crises in childhood were stronger for women than for men. The effect of nutritional defects on fetal growth — leading to impaired health in later life was tested using an HSN cohort that had experienced the potato blight of the 1840s. Van den Berg, Lindeboom, and Portrait (2013) discovered that men exposed to severe famine during pregnancy (at least four months) and directly after birth have a significant lower residual life expectancy at age 50 than others, but not at earlier ages. The effect was not found for women, or for men experiencing the crisis when they were slightly older (also [Lindeboom, Portrait, & van den Berg, 2010](#); [Popławska, 2015](#)). Schellekens and van Poppel (2016) use aggregate mortality rates and height as indicators of early life conditions and estimate that improvements have contributed more than five years, or about a third, to the rise in life expectancy of females at age 30 between cohorts born in the periods 1812–29 and 1910–21. And such improvements contributed almost three years, or more than a quarter, to the rise in life expectancy of males at age thirty.

Research has also focused on 'intermediate' events in the life course which could exacerbate or weaken effects from early life (on the role of marriage, see [van den Berg and Gupta \(2015\)](#); also [Gupta \(2010\)](#)). Effects from mid-life experiences on health in later life were also studied. Kaptijn et al. (2015) test the hypothesis of a trade-off between women's fertility and their longevity. Fertile women may have shorter lives, e.g. through maternal depletion. Such a trade-off was found, but only for the oldest cohorts. Women with few, or with many children, lived shorter. Possibly, improved wealth and health conditions explain the disappearance of the trade-off in the younger cohorts. Stressful periods can also lead to impaired health. Alter, Dribe, and van Poppel (2007) showed that widows, especially those with large families, lived shorter than other women. Migration might also be such a stressful experience. However, Puschmann (2015), Puschmann, Donrovich, Grönberg, Dekeyser, and Matthijs (2016) and Puschmann, Donrovich, and Matthijs (2017), using among others HSN data on Rotterdam, demonstrated the existence of the 'healthy migrant' effect; migrants were already positively selected and actually had lower mortality risks than the natives. This proved especially true for Zeeland and Brabant migrants to Rotterdam. However, they also detected specific vulnerable groups with higher mortality risks, such as Italian migrants.

#### 4.4 FAMILY AND SURVIVAL

In line with the already discussed interest in 'cooperative breeding' and related hypotheses stemming from evolutionary biology, researchers have looked for the effects of living-in kin on child survival. Clearly the presence of parents, especially the mother is of paramount importance (van Poppel & van Gaalen, 2008b). Kok, Vandezande, and Mandemakers (2011) showed that living-in kin could have very beneficial effects. The positive effects of the presence of grandparents could compensate for the loss of the father or mother. Uncles and aunts as well played a positive, altruistic role. Apparently, a family setting which was optimal for infants' and children's well-being included parents, grandparents, (unmarried) uncles and aunts, possibly servants, boarders and lodgers. However, the family should not have too many young children, as the infant would have to compete with them (also Vandezande, Kok, & Mandemakers, 2011).

Exploring this 'competition' further, Riswick (2018) used the HSN to study how infant and child mortality was affected by the interaction between resource dilution and the specific family type prevailing in different regions in the Netherlands. He shows, firstly, that the number and gender of siblings played an important role in determining child mortality, but were less significant in determining infant mortality. After the age of one, boys experienced more competition from their brothers. Riswick surmises that this effect is related to their assigned roles within the household and on the family farm. A similar effect was found for girls in the nuclear family region in the northwest. Here, fewer children were needed to work inside or outside the household, and girls and boys may even have been interchangeable.

#### 4.5 RELIGION AND THE LIFE COURSE

The Dutch population registers are among the few that record religious denomination. And they did so quite meticulously: more than a hundred different denominations can be found in the Dutch registers. This offers the rare opportunity to study how religion can affect demographic behavior and life course transitions. In principle, it is possible that such effects were caused by specific religious norms and prescriptions, coupled with effective sanction mechanisms. However, McQuillan (2004) has argued that another condition is necessary, which is the identification of people with their clergy. The latter could differ from country to country, and was often due to political reasons. But apart from religious norms, sanctions and identification, belonging to a church *community* may have an effect in itself, e.g. through the support derived from the community. Observed differentials by religion may also stem from differences in social characteristics of the members. And finally, the observed connections may not derive from *individual* beliefs and associated behavior, but from an underlying effect of the presence of specific churches on the environment (Kok, 2017). Kok compared the outcomes on the religion variables in regressions models on twenty different life course transitions, ranging from leaving home to old age mortality. Differences in moral prescriptions and sanctions between the Churches, but, perhaps even more important, in 'mentality' (e.g. fatalism or openness to the outside world) led to marked variation in behavior, after controlling as much as possible for social characteristics.

One of the HSN subprojects has been an oversample of Amsterdam Jews. This material has been used to study how (during 1880–1940) assimilation into Dutch society at large changed their household structure (Tammes & Van Poppel, 2012). Furthermore, their life courses were studied to understand how and why Jews decided to leave their faith, and to become either unaffiliated or to convert to Christianity. Individual backgrounds were combined to period effects (changes in the level of assimilation of Jews through intermarriage, the rise in atheism, and anti-Semitism), and the outcomes suggest different pathways to either conversion to Christianity or to simply abandon the faith (also Tammes, 2012b; Tammes & Scholten, 2017). Tammes (2011) also used the data to study residential segregation of Jews before the War. The rise of secularization among the HSN population was the subject of a study by Knippenberg and de Vos (2008).

Jews and Roman Catholics had been subject to official discrimination before 1800. Van Poppel, Liefbroer, and Schellekens (2003) showed, on the basis of a specifically constructed HSN subset, that this indeed affected their (inherited) class positions (in the city of The Hague), but also that intergenerational social mobility patterns in the 19th century did not differ across the denominations. In other words, social disparities between religions were 'path dependent' and not the result of 'new' obstacles.

## 4.6 SOCIAL NETWORKS

Register data as such give very little clues about the social networks of the RPs. At best, we know who, for a longer or shorter time, lived in one household. With respect to kin networks, this lacuna can be dealt with by combining HSN to the extended kin traced through the LINKS matching system, see also below. But friends, colleagues and acquaintances remain out of view. However, the civil records, which are also part of the HSN database contain witnesses at birth, marriage and death. Especially those present at the marriage are of interest, as the relation to the bride and groom are specified. Although witnesses can be simply municipal clerks, or even witnesses from a just completed wedding asked to stay on for the next (Kok, Adams, Ericsson, & Moch, 2002), generally the data give a good insight in who were considered close and important. Van Poppel and Schoonheim (2005) used the witnesses to study the networks of Jews, in contrast to those of other religions. They observed a much stronger familial involvement with the marriage ceremony among Jews, a phenomenon that was observed in all social classes. Also, the network of Jews included people coming from a much wider area than was the case among Dutch Reformed and Catholics. Again, these differences remained when social class of the groom was controlled (also Tammes, 2012a).

Bras (2011) uses the marriage witnesses in HSN to study a transition in the meaning of kinship. During the 19th and 20th centuries, an intensification of family relations took place (see also Bras, 2014a). During the period 1830–1950, lateral kin (siblings, siblings-in-law and cousins) were increasingly selected as marriage witnesses, at the expense of professional witnesses and patronage relations. The process started in urban higher and middle classes and among farmers, but cultural diffusion took place whereby the choice for brothers(-in-law) and cousins increasingly spread to other social groups. Also, familial age peers were particularly selected when the mother (of the bride) was still alive and when the bride was literate, pointing to the important role of women in cultivating family relations (see also Zijdemans & Corten, 2012).

## 4.7 BIOLOGICAL STANDARD OF LIVING

Economic history, as well as other disciplines, shows an increasing interest in new indicators to measure well-being, in terms of determinants, long-term developments, societal variation and inequality. Prominent among these new indicators are human heights, as they reflect (net) nutrition, health and workload (Floud, Fogel, Harris, & Hong, 2011). HSN and the development of the LINKS dataset, in combination with an integrated index to the municipal militia records (thanks to a crowdsourcing project) enabled a unique research design: the linking of complete life courses to early adult heights of RPs as well as of their (male) kin. The project *Giants of the Modern World* aims to understand what changes in Dutch society (e.g. in family size, access to high quality food, economic growth and integration) were responsible for the unique position of the Dutch as the tallest in the world. The first results confirm that in the second half of the 19th century the coastal provinces and large cities recovered from the economic stagnation that had plagued them since the late 18th century, but national convergence was still not visible by 1940 (Tassenaar, 2019).

The HSN makes it possible to study the effect of (changing) early life conditions on stature, because all changes in household composition are known from birth of the RP onwards. Quanjer and Kok (2019) experimented with a consumer/producer ratio and showed that it captures resource dilution effects in households more effectively than e.g. sibship size. They also demonstrated that children suffered more from the death of the mother than from the death of the father (see also 3.3.2.) and that boys had an advantage over girls when it came to food distribution in households (at least they grew taller when they had female siblings compared to male siblings). Finally, they conclude that the early introduction in the Netherlands of the homemaker-breadwinner model might be an important clue to understanding Dutch giantism.

Stature not only reflects adverse or beneficial early life conditions, it can also have an independent effect on one's life chances, e.g. by its relation to earning capacity. To unravel direct and indirect effects on later life outcomes, one needs 'mediation' models. Thompson, Lindeboom, and Portrait (2019) used such models to relate economic crises in childhood (the Potato Famine of 1817) to socio-economic status in early adulthood, mediated by height. So far, no mediation effect was found.

## 5 META RESEARCH ON DATABASES, DATA STRUCTURES AND DATA

### 5.1 DATA STRUCTURE

From the start the HSN worked with custom made data entry programs. When the database became larger, we gained a better understanding of how to organize the data. Although numerous studies existed detailing the process of historical data entry and the difficulties of data interpretation, how the database itself should be organized remained quite implicit. In May 2001, the HSN organized a workshop in which all major database administrators were invited to comment on the sketches and ideas that existed within the HSN. This workshop was very well attended and resulted in a comprehensive list of best practices for the creation of large databases on historical populations (Mandemakers & Dillon, 2004). Most important in these guidelines was the systematic distinction between the data entry part of the database, the database itself with unstandardized, standardized, enriched and linked data into life courses and the part with all the different releases that have been realized during the history of the database.

Building life courses from dynamic data such as data from population registers is not an easy and straightforward job. The way the HSN constructed the life courses out of the raw data was first systematically described by Mandemakers (2006b). During this process several decisions about the structure of the dataset had to be taken that were not always in the interest of all researchers. It turned out that this was not a typical HSN problem but a more general one for all historical databases. And there was also the question of how to enlarge the potential use of all these databases. Several workshops were organized to tackle this problem, a first one in Montréal (2003), a second one in Amsterdam (2006) and a third one in Ann Arbor (2008). Although all participants in the meetings recognized the complex nature of their longitudinal databases, it was the general feeling that comparative research was possible if certain conditions were set. It was agreed, firstly, that standardization would be the basis for comparative research. And secondly, that a standard structure was necessary to create possibilities of data conversion from different datasets into common datasets for analysis. This resulted in a model for data sharing, which became known as the Intermediate Data Structure (IDS; Alter & Mandemakers, 2014; Alter, Mandemakers, & Guttman, 2009). Several databases have been converted into the IDS-structure. Major developments are the integration of four Swedish databases into one IDS data structure (Swedpop project, <https://www.vr.se/english/mandates/research-infrastructure/find-research-infrastructure/list/2018-10-18-swedpop---swedish-population-databases-for-research.html>). The HSN is working to integrate the HSN register data with the civil certificates and the LINKS dataset. LINKS will provide more data on already included relatives and will also extend the number of relatives as far as third and fourth degrees. So-called extraction software to convert IDS-data into datasets for analysis have been developed by Luciana Quaranta (2015, 2016, 2018).

Within the context of the HSN the existing knowledge was also used to develop data structures of data entry or integration for data from former colonies of the Netherlands. Ideas to develop HSN like databases for Suriname and Dutch East Indies, especially Java, were developed and tested (Bosschaart & Lelieveld, 2016; van Bruggen, Bulten, & Maas, 2017). Also, a data structure was developed for the so-called *thombos*, a quite complicated population and land register of Dutch colonial Ceylon from the second part of the 18th century (van den Belt, Kok, & Mandemakers, 2011). The inaugural lecture of Mandemakers (2009) stressed the importance of standard formats and the cooperation of specific scientists from different academic fields, statisticians and methodologists to do cutting edge research.

The first initiative to develop the Intermediate Data Structure was supported by the Dutch Research Council (NWO; the Humanities program for internationalization). This initiative was strengthened by the so-called *European Historical Population Samples Network* funded by the European Science Foundation (ESF). This program ran from July 2011 to December 2016 and brought together in one program databases from about twenty European countries, whereas databases from outside the EU participated as much as possible. The program was chaired by the HSN and concentrated on several tasks: documenting of existing databases, promoting of the IDS and developing software that ran on the basis of the IDS structure and organizing summer schools. All documentation was concentrated on the website, including the launch of a journal to publish methodological and substantive research articles, especially if the used datasets were based on the IDS, see further <https://ehps-net.eu/>. A follow up of the program was *Methodologies and Data mining techniques for the analysis of Big Data based on Longitudinal*

*Population and Epidemiological Registers (LONGPOP)*. It is a project within the framework of the Marie Skłodowska-Curie Innovative Training Network within the Horizon 2020 Programme of the European Commission and is led by Diego Ramiro (<http://longpop-itn.eu/>).

A new development in data formats is the introduction of so-called linked open data structures (LOD). By way of the semantic web it becomes possible to connect data from databases all over the world and to create 'new' datasets (Meroño-Peñuela et al., 2015). Initiated by DANS, the HSN and Free University Amsterdam cooperated in the CEDAR project to develop a LOD structure for the aggregate data of the Dutch censuses from 1830 till 1948. This resulted in a methodological description of workflows to convert source data into the new data structures in a systematic and repeatable way (Ashkpour, 2019; Ashkpour, Mandemakers, & Boonstra, 2016; Ashkpour, Meroño-Peñuela, & Mandemakers, 2015).

## 5.2 STANDARDIZING AND CODING DATA

HSN occupational data as such were intensively used during the construction of the Historical International Classification of Occupations, the so-called HISCO code (van Leeuwen, Maas, & Miles, 2002, 2004). This was not that remarkable since three researchers from the HSN steering committee, Marco van Leeuwen, Ineke Maas and Onno Boonstra, were strongly involved in this enterprise (Maas, 1998). The introduction of the HISCO code, by Theo Engelen and Hans Hillebrand, lamented the situation that in the Netherlands 'each researcher constructed his/her own classification scheme' (Engelen & Hillebrand, 1985, p. 255).

The HISCO code itself is not suitable for doing analytic research, but it forms the basis for stratification schemes according to social class or social status. Using characteristics of the occupational titles according to distinctions like manual or non-manual character, higher, medium, lower or no schooling, supervising or not and being part of the primary sector, Maas and van Leeuwen constructed a decision scheme based on HISCO resulting in twelve social classes (HISCLASS). Van de Putte and Miles (2005) constructed SOCPO which is a scheme of five status groups based on the principal of social power originating from both economic and cultural resources. Economic power is derived from factors such as self-employment, skill, and authority (command). Sources of cultural power are non-manual versus manual occupations and nobility and prestige titles. An interval scale called HISCAM created a scale from 0 to 100, ranking the HISCO codes on the basis of the occupations of the participants in marriage certificates (Lambert, Zijdeman, van Leeuwen, Maas, & Prandy, 2013; Zijdeman & Lambert, 2010).

During the last twenty years HISCO has become the basis for the classification and ranking of occupational titles for almost all research with HSN data. The HSN has also taken the lead in standardizing and classifying occupational titles into HISCO, HISCLASS, HISCAM and other systems. The latest release (Mandemakers et al., 2018) includes 134,964 different occupational titles found in the sources used by the HSN and LINKS datasets.

Besides the coding of occupational titles, standardization schemes have been developed for all major data types that are included in the HSN and LINKS datasets such as locations, religion, age, civil status, types of certificates, sex, first names and last names. Most important is the location file with standardized and spatio-temporal coded values of 7,925 location names (hamlets, villages, towns, cities and municipalities). This dataset was released by Huijsmans (2013) including the municipality to which they belong, X and Y coordinates as well as longitude/latitude coordinates within the framework of the LINKS project. And of course, the release also includes the AMCO code developed by van der Meer and Boonstra (2006) which is the way to link with all other datasets in the field of locations.

## 5.3 RECORD LINKAGE

In essence the LINKS project is no more than software for record linkage and standardization in which the indices of names from the Dutch civil registration are used to reconstruct pedigrees and complete families for the Netherlands in the 19th and first half of the 20th century. No wonder that a lot of effort went into improving record linkage procedures. The first record linkage on marriage certificates was done by Oosten (2008) who developed artificial intelligence to match equal but apparently different names such as Guillaume and William. The basic idea of his algorithm was that two of the same names will never appear in first names that are a combination of different first names (like 'Cornelis Albertus'). The idea was that if a positive match was only dependent on two completely different first names

and these two names did not appear together in combination of first names than the names probably would be synonyms of each other (if not they should appear together in combined first names).

In his thesis research Schraagen (2014) tested old and developed new techniques on record linkage with the context of the LINKS project. Together with Bloothoofdt he further developed the basic ideas of Oosten by learning name variants from matches that on the one hand have a high level of precision, but on the other hand lacked one element necessary to approve for a match (Bloothoofdt & Schraagen, 2015). Schraagen experimented with different and innovative record linkage techniques, such as using domain based information on the structure of families resulting in more valid evidence for proposed links (Schraagen & Kusters, 2014), using graph theory to predict if a link is missing because of failed record linkage or because of missing records (Schraagen & Hoogeboom, 2011) and developing methods to compare datasets which are structured differently for events (from certificates) or for individuals (from genealogies; Schraagen & Huijsmans, 2013). Within the LINKS project the emphasis gradually changed from record linkage as a technique to the reconstructions of populations as a result. In 2014 LINKS organized a conference with the theme 'Population Reconstruction'. Several topics were presented and discussed on themes as standardization of names and locations, record linkage and population reconstructions (Bloothoofdt, Christen, Mandemakers, & Schraagen, 2015).

## 5.4 STUDIES ABOUT DATA AND DATA QUALITY

Throughout the years, researchers within the HSN domain learned to be critical of the historical data they were using. Using marriage certificates for social mobility, for example, will be associated with various problems. Firstly, the occupational titles of the bride and especially the mothers are underreported, secondly no occupational titles are included when a parent is not alive anymore and thirdly the occupational titles of father and son are only associable at different age levels. Delger and Kok (1998) researched this source of bias by comparing mobility tables for the same HSN RPs based on marriage certificates and based on data from the population registers. The last ones include more fathers with occupations and are also better comparable because the occupations can be compared at the same age levels. The comparison made clear that mobility is underestimated when only the marriage certificates are used as data source. The sons whose fathers had already deceased when they married were characterized by relatively more downward mobility. Walhout and van Poppel (2003) systematically researched when the brides presented an occupational title in the marriage certificate or not. And the sampled character of the birth certificates of the HSN made it possible to do research on the nature of the witnesses in the certificates. Were they family, or so-called professional witnesses hanging around municipality halls having not much to do (Mandemakers, 2017)?

From the onomastic perspective names were not only used as a way to link certificates, but as such they were studied as well. Bloothoofdt, Mandemakers, Brouwer, and Brouwer (2012) overviewed the way family names and first names are available for research from 1811 till the present. Doreen Gerritzen used the first names of the HSN basic sample to make popularity polls of first names during the 19th and 20th centuries for several provinces (Gerritzen, 1998, 2001). For the introduction, development and socio-cultural imbedding of naming by way of combined first names was researched from 1760 onwards till the present (Bloothoofdt & Onland, 2016).

Register data, although superior to many other historical sources on populations, obviously come with biases and lacunae of their own (Kok, 2006b). Biases may relate to the observations themselves, to the used administrative categories, to the information provided, and to the fixation on addresses of the population registers. Given the richness of Dutch administrative sources some of the mentioned problems could be solved.

As to observations: the thorough system of civil registration, introduced in the wake of the French occupation, does not leave much room for misidentification or 'disappearance' of RPs. A small number of RPs could not be traced to the municipalities where they lived, because their mother was either single (which frequently implied she was sent to an institution or hospital to give birth), a vagrant, or a skipper's wife. In those cases, the child was not born in the place where the mother officially resided. More serious are (temporary) gaps in the observations across the life course, when persons failed to officially declare their move to another locality, or simply because registers did not survive the Second World War or flooding. The system of Personal Cards, which started in 1940, allowed HSN staff to pick up those 'disappeared' persons in 1940, and trace their moves backward in time. A systematic study of these so-called cold cases showed that about a third was traceable, but the rest was not mostly because of the destruction of population registers by war actions or flooding (Haarman, 2018). However, in

many life courses of RPs, shorter or longer periods are missing. Also, as has been said before, the sample only covers Dutch natives, not immigrants. Finally, RPs in the youngest cohort may still be alive today. This implied that no Personal Cards were available, and the end of observation, at best, had to be put at 1940. In particular early studies on mortality based on the first HSN releases were beset with this problem. Van Poppel and van Gaalen (2008a) surmise that mortality in the youngest cohorts was overestimated because of this problem.

Interpreting life courses and demographic behavior with HSN implies having to rely to a large extent on the administrative categories devised in the 19th century. Although most of them are rather straightforward, we do not always know how far they reflect historical reality. An administrative head of the household may not have been a person of real standing at the time. Moreover, as changes in headship are not always recorded, HSN assigns headship status according to an algorithm which may not always be accurate (Mandemakers, 2006b). Sometimes, administrative procedures were ideologically driven — as in the case of Rotterdam after about 1910. Here, the population registers were restructured to fit the ideal of the nuclear family. Not only were living-in kin assigned to cards of their own, single mothers and their children were also separated administratively. Thus, we need to look through 'administrative' lenses to critically assess the value of all information from the registers. For instance, official moves to another address were recorded, but lengthy stays from home (sailors, seasonal workers) remain invisible. Finally, and obviously, the official religious denomination is not a proper indicator for one's religiosity.

The quality of information in the population registers may vary. Firstly, in contrast to civil records, the information is given by informants without the official approval of witnesses. Thus, especially in larger places where they were unknown to the officials, persons may have given false information. Examples are cohabiting couples who reported the female partner as 'housekeeper', or cases of single motherhood where the child is ascribed to the grandmother. Secondly, individual changes in occupation and religious affiliation were generally only recorded at the time of the decennial update, or when people migrated. Thus, to some extent the socio-economic profile of an RP depends on geographic mobility. Information in the registers on heads of household may be of better quality than of his or her dependents as occupations of women and children were often not recorded. Therefore, the official occupation of the head may give insufficient information on income and status of the entire household. Thirdly, the SES information contained in the registers is limited to occupations, which, although providing important information, sometimes is too broad ('merchant', 'farmer', 'worker') to indicate the income level and status of the family. Kok, Mandemakers, and Damsma (2010) showed that adding income level for the farmers led to a much more nuanced picture: their income level determined strongly the timing and incidence of the marriage of their children. Also, persons without occupations may be unemployed, retired, or actually living handsomely on non-labor income without us knowing. Finally, missing information may actually lead to biased conclusions. For instance, the strong association Van Bavel and Kok (2010a) found between religiously mixed marriages and childlessness may have resulted from an (unobserved) effect of education, if mixed couples had received relatively high education.

As to the fixation on addresses: in the research design of HSN only the Research Persons are traced in the population registers. This implies that other individuals and kin are out of sight when they don't live at the same address as the RP anymore (Mandemakers, 2006b). For instance, very close kin may actually live next door (or even in the same house when the address was split), without the researcher being aware. The administration also 'fixes' persons to a specific household. The current situation of children from broken families spending time in different households may have occurred — in other forms — in the past as well (van Poppel, Schenk, & van Gaalen, 2013).

## 6 PROSPECTS

Recently, many historical demographers have joined forces and put forward their ideas for the future of the field (Matthijs, Hin, Kok, & Matsuo, 2016). How does the HSN fit in with these visions? Several scholars have argued for an extension backwards in time and in place, covering non-European areas. In principle, the datasets of the HSN and LINKS could go further backwards into the 18th and 17th century by sampling and linking parish records, but this is not anticipated in the near future. As to

space, a pilot study is undertaken to create a Historical Sample of the Dutch East Indies (van Bruggen, Bulten, & Maas, 2017). Also, the HSN is involved in the construction of a Historical Database of Suriname (van Galen & Hassankhan, 2018).

Some researchers (e.g. Kurosu, 2016) plead for more comparative research, following the lead of the Eurasia Project in which high-quality datasets on rural populations from several countries were compared. Indeed, this is an exciting prospect, as more life course datasets, genealogies, etc. are becoming available. HSN has already been involved in several comparative projects. Among the most detailed population registers in the world are the Taiwanese household records created by the Japanese colonial government. Since about 2000, HSN has been used to contrast Dutch family formation processes and mortality with the Taiwanese experience (Kok, Yang, & Hsieh, 2006; Lin, 2011; Riswick, 2020; Shephard, Kok, & Hsieh, 2006; Shephard, Pan, Kok, Engel, Engelen, & Brown, 2006; Shephard, Schoonheim, Chang, & Kok, 2011). Other comparative projects revolve around fertility reactions to child mortality (Reher, Sandström, Sanz-Gimeno, & van Poppel, 2017) or the intergenerational transmission of infant mortality (Quaranta & Sommerseth, 2018).

Others have advocated a more qualitative approach to demographic behavior, and have pleaded for more attention to culture and to the historical context in which, e.g. occupations and causes of death were recorded (Brandström, 2016). Indeed, a continuous effort is needed to a) understand the data we have collected in terms of their origin, specificity and local variation, b) to improve documentation and ensure awareness among users of the contextual nature of the data and c) to continue adding variables. For instance, variables on income, schooling and health can help to avoid an overreliance on occupations as the major indicator of SES. As for culture, new research into religiosity (e.g. by going into church archives or studying naming practices) can complement the already extensive findings on the effects of denomination on life course transitions. Another direction is adding data on local popular beliefs and relating them to demographic behavior. An example is research by Hilde Bras who linked a survey of the Meertens Institute (dedicated to the study of dialect and folklore) on beliefs regarding isolation of women in childbed and enchantment of children to local variations in fertility (Bras, 2014b).

Smith, Hanson, and Mineau (2016) argued that historical demography should link up with the trend towards Big Data, which would increase its relevance for demography at large. They point at the rapidly increasing possibilities to use massive numbers of digitized family trees. As soon as the HSN is linked to LINKS, the problem of 'atomization' by focusing on individuals devoid of their (kin) networks will be solved. It will become possible to trace ancestors, extended kin and descendants of HSN RPs and to study how the careers, migrations, family formation and health was affected by intergenerational transmission (see Bras, Van Bavel, & Mandemakers, 2013; Murphy, 2013) and by kin proximity, support and social status. Tracing descendants of HSN RPs will also stimulate collaboration between historical demographers and genetics (see also Hobcraft, 2006; Larmuseau, Van Gestelen, van Oven, & Decorte, 2013). In fact, descendants of HSN RPs are already being traced in order to link their DNA samples to the health status of their ancestors (project *Genes, Germs and Resources*, see <https://iisg.amsterdam/en/hsn/projects/ggr>).

By using remote access linkage to archives containing individual causes of death (after 1936), the HSN will also be useful for health-related research in the future. Another important step is adding the height of HSN men at age 19, determined at the physical examination for military service (research project *Giants of the Modern World*). HSN can also expand its mission by offering more information on the environment in which people lived: soil and water condition of the area, precipitation and temperature, population density, transport and communication infrastructure, presence of industry, and local causes of death (Ekamper, 2019). This would create the setting for a real multilevel analysis of demographic behavior. Another envisaged development is the connecting of the younger cohorts of the HSN with the Social Statistical Datasets (SSD), curated by Statistics Netherlands. The SSD integrates register and survey data on all inhabitants of the Netherlands since the 1990s, containing important socio-economic, demographic, health, crime, income and wealth variables of the complete population of the Netherlands. It includes day-by-day information on co-residence, marriage and parenthood and detailed information on parents, children, siblings, partners, colleagues, companies and communities and neighbors (Bakker, van Rooijen, & van Toor, 2014). The HSN will be connected by way of the second generations of the HSN research persons, born from 1900 onwards.

It can be foreseen that future generations of HSN users will work in even more interdisciplinary settings, e.g. by working with evolutionary biologists. But probably the largest challenge will be to translate

the findings of HSN studies to the larger historical processes. What does all the variation they find at the individual or household level imply for the trends in society observed at the aggregate level? The methodological bridge between the micro and macro levels can possibly be found in *agent-based modelling*. This technique goes beyond classic micro simulation (creating synthetic populations from demographic characteristics and rates culled from e.g. registers) by letting the simulated persons learn from their past, react to others, influence their surrounding and be influenced by their environment. Much more social variation can be built in such models than was possible in the past. In this way, different pathways of, for example fertility decline through diffusion of innovative ideas, can be tested (Billari & Prskawetz, 2003; Courgeau, Bijak, Franck, & Silverman, 2016; for examples of modelling the fertility decline, see Klüsener, Scalone, & Dribe, 2016; Nomes, Grow, & Van Bavel, 2019).

## 7 EVALUATION

We began this essay with the question on the 'qualitative impact' of the HSN. Has HSN changed our knowledge and perception of the history of Dutch society, and in what ways? What is HSN's contribution to social science history in general and to the specialist field of (historical) database construction? A synthesizing social and demographic history of the Netherlands based on HSN still has to be written. As the foregoing has shown, many new insights are scattered across numerous articles. HSN-based research has shown the 'micro-effects', that is the implications for individuals and families, of 'macro' characteristics of Dutch society in 19th and early 20th centuries. Such characteristics are: a relatively wealthy country with a generous poor relief system; a late-comer in industrialization; an early adopter of the breadwinner-homemaker family model; and a country 'pillarized' along religious fault lines.

Although particular western areas of the Netherlands suffered economic hardship in the early half of the 19th century, we do not see the same strong effects of food price fluctuations as in other European countries. Apart from market integration, it seems that the poor relief system which was built up in more prosperous eras managed to mitigate effects of bad times and the restructuring of the economy. For instance, social class differentials in adult mortality were limited and widows seem to have been shielded against the loss of the breadwinner. Effects of industrialization (through the changing occupational structure, the valuation of achievement and female factory employment) have been studied on issues such as migration, social mobility and fertility decline. Clear effects were not found, probably because industrialization came rather late and coincided with increasing real wages and expansion of education. As for migration, the loss of opportunities in the countryside led to migration flows into the service sectors of the large cities, to emigration, and to a limited extent to the new industrial areas.

For many centuries, the Netherlands have been a typical 'nuclear family' country, apart from the border area with Germany. Its relative wealth made it also possible to implement the ideal of the homemaker model, with women withdrawing from the labor market and concentrating on the household. HSN had allowed to trace this development and to study the effects, e.g. on the decline of infant mortality and increased stature (read: health) of children. Finally, the Netherlands were characterized by increasing competition among religious denominations, in which (sexual) morality played an important role. In particular Roman Catholic and neo-Calvinist foremen endeavored to discipline their flocks and to moralize the country at large. HSN research shows how this played out in diverging sexual and marital behavior among different religious groups.

A typical feature of HSN research is its international character. Thus, many studies were comparative and/or oriented at an international audience. For example, several researches on fertility behavior (e.g. response to declining infant mortality, spacing of births) were done in comparison with other countries. Complete residential histories covering an entire country are extremely rare, and as such HSN has been used as a 'benchmark' to see where gaps and biases could occur in other data collections. Examples are the comparison of migration trajectories reconstructed from population registers (HSN) and those from genealogies (Adams, Kasakoff, & Kok, 2002) and using HSN to study the attrition process and selection mechanisms before boys could be measured at conscription age (Quanjer & Kok, 2020). Also, the Netherlands were presented as a 'laboratory' to study specific behavior that could not (or not yet) be studied elsewhere. A case in point is the relative impact of socio-economic status, household

composition and religion on a host of life course transition and life events. HSN researchers eagerly joined international debates and innovations in methods and research topics. Examples are studies on (household strategies behind) leaving home (e.g. [Bras, 2002](#)), on 'helpers at the nest' and their effect on reproduction (e.g. [Rotering & Bras, 2015](#)) on extended kin and social networks (through witnesses) and their role on social mobility and diffusion of birth control (see [Bras, 2014a](#)), on life course approaches to anthropometric history (e.g. [Quanjer & Kok, 2019](#)) and so on.

A major spin-off of HSN is LINKS, which has quickly become an important player in the field of studies on intergenerational effects in social mobility and health. Examples are the recent studies on familial clustering of infant death and longevity. Social and life science struggle with disentangling environmental and heritable components of behavior and health, and LINKS promises to become a crucial tool, especially when the entire country is integrated in LINKS and when it can be enriched with more variables.

HSN and LINKS are much more than just datasets. The HSN is a center of expertise in collecting, harmonizing and documenting data from a variety of historical sources, in record linkage, and in creating standards for occupations and locations. Moreover, the HSN network of researchers have set an example for Dutch scholarly work (at least, in the humanities) by collaborating in interdisciplinary teams and by sharing data and tools.

## 8 EPILOGUE

When thirty-odd years ago the HSN was conceived, we anticipated that it was going to be a long-lasting undertaking. The data were available in Open Access from the start which ensured interest from many researchers both from the Netherlands and abroad. The database — or more precisely data warehouse — was designed to expand. By enlarging the geographical scope and the period covered and by constantly adding variables allowing for detailed life courses, the HSN ensured a growing (and even insatiable) interest from — by now — several generations of researchers. Furthermore, the many additional projects designed for specific researchers (e.g. an oversampling of city, the addition of siblings of RPs, an entire second and third generation of selected RPs) were added to the warehouse after an initial embargo period in order to be reused in further research.

What we could not anticipate around 1990 were the technological advances realized in the ICT sector in the ensuing years. At the time, we were happy with the 3.5 inch diskettes on which our employees in the archives could store the data. Now, we can effortlessly (in term of memory storage and processor speed) link HSN to other databases (e.g. using the common IDS format) or to enrich the data in the Cloud by record linkage in Open Linked Data. Obviously, the speed and scale of actual research has gone beyond our wildest dreams.

Another matter we could not foresee at the start was the sheer expansion of research questions, which of course stems from the dynamism of science itself. For instance, great strides in the sociology of the life course were made from the 1990s onwards. Also, emerging intersections between evolutionary biology, biodemography and historical demography lead to new questions being researched with HSN and similar databases. Another example is the strong rise of genealogical demography and health studies related to the revolution in genetics.

What we can foresee is that HSN and LINKS will remain at the core of an expanding ecology of national and international historical micro-level databases, provided researchers and funding agencies go on investing time, money and energy in keeping the datasets up to standards and continue catering to the ever changing needs of the research community.

## ACKNOWLEDGEMENTS

We are grateful to Frouke Hansum who corrected the English text and Marja Koster who systematized our provisional reference list into the required format for this journal.

The article is written within the context of the project 'Methodologies and Data mining techniques for the analysis of Big Data based on Longitudinal Population and Epidemiological Registers' (LONGPOP). The project is funded by the EU Horizon 2020 research and innovation programme (under the Marie Skłodowska-Curie Grant Agreement 676060). DISCLAIMER: the results published in this website reflect only the author(s)'s view and the Research Executive Agency is not responsible for any use that may be made of the information they contain.

## REFERENCES

- Adams, J. W., Kasakoff, A., & Kok, J. (2002). Migration over the life course in XIXth century Netherlands and the American North. A comparative analysis based on genealogies and population registers. *Annales de Démographie Historique*, 104(2), 5–27. doi: [10.3917/adh.104.0005](https://doi.org/10.3917/adh.104.0005)
- Alter, G., Dribe, M. & van Poppel, F. (2007). Widowhood, family size and post-reproductive mortality: A comparative analysis of three populations in nineteenth-century Europe. *Demography*, 44(4), 785–806. doi: [10.1353/dem.2007.0037](https://doi.org/10.1353/dem.2007.0037)
- Alter, G., & Mandemakers, K. (2014). The Intermediate Data Structure (IDS) for longitudinal historical microdata, version 4. *Historical Life Course Studies*, 1, 1–26. Retrieved from <http://hdl.handle.net/10622/23526343-2014-0001?locatt=view:master>
- Alter, G., & Mandemakers, K., & Gutmann, M. (2009). Defining and distributing longitudinal historical data in a general way through an intermediate structure. *Historical Social Research*, 34(1), 78–114. doi: [10.12759/hsr.34.2009.3.78-114](https://doi.org/10.12759/hsr.34.2009.3.78-114)
- Ashkpour, A. (2019). *Theory and practice of historical census data harmonization: The Dutch historical census use case: A flexible, structured and accountable approach using linked data technology* (PhD thesis Erasmus University Rotterdam). Retrieved from <http://hdl.handle.net/1765/114388>
- Ashkpour, A., Mandemakers, K., & Boonstra, O. (2016). Source oriented harmonization of aggregate historical census data: A flexible and accountable approach in RDF. *Historical Social Research*, 41(4), 291–321. doi: [10.12759/hsr.41.2016.4.291-321](https://doi.org/10.12759/hsr.41.2016.4.291-321)
- Ashkpour, A., Meroño-Peñuela, A., & Mandemakers, K. (2015). The aggregate Dutch historical censuses. Harmonization and RDF. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 48(4), 230–245. doi: [10.1080/01615440.2015.1026009](https://doi.org/10.1080/01615440.2015.1026009)
- Bakker, B.F.M., van Rooijen, J, & van Toor, L. (2014). The system of social statistical datasets of Statistics Netherlands: An integral approach to the production of register-based social statistics. *Statistical Journal of the IAOS*, 30(4), 411–424. doi: [10.3233/SJI-140803](https://doi.org/10.3233/SJI-140803)
- Barker, D. J. P. (1992). *The fetal and infant origins of adult disease*. London: BMJ Books.
- Bengtsson, T., & van Poppel, F. (2011). Socioeconomic inequalities in death from past to present: An introduction. *Explorations in Economic History*, 48(3), 343–356. doi: [10.1016/j.eeh.2011.05.004](https://doi.org/10.1016/j.eeh.2011.05.004)
- Billari, F. C., & Prskawetz, A. (Eds.). (2003). *Agent-based computational demography. Using simulation to improve our understanding of demographic behaviour*. Berlin-New York, etc.: Springer.
- Blau, P. M., & Duncan, O. D. (1967). *The American Occupational Structure*. New York: Wiley.
- Bloothoof, G., Christen, P., Mandemakers, K., & Schraagen, M. (Eds.). (2015). *Population reconstruction*. Cham Heidelberg New York Dordrecht London: Springer. doi: [10.1007/978-3-319-19884-2](https://doi.org/10.1007/978-3-319-19884-2)
- Bloothoof, G., Mandemakers, K., Brouwer, L., & Brouwer, M. (2012). Data mining in the Dutch (historical) civil registration (1811–present). In P. Darlu, G. Bloothoof, A. Boattini, L. Brouwer, M. Brouwer, G. Brunet, G., ... F. Manni, *The family name as socio-cultural feature and genetic metaphor: From concepts to methods*, *Human Biology*, 84(2), 177–184. Available from <http://digitalcommons.wayne.edu/humbiol/vol84/iss2/5>
- Bloothoof, G., & Onland, D. (2016). Multiple first names in the Netherlands (1760–2014). *Names*, 64(1), 3–18. doi: [10.1080/00277738.2016.1118860](https://doi.org/10.1080/00277738.2016.1118860)
- Bloothoof, G., & Schraagen, M. (2015). Learning name variants from inexact high-confidence matches. In G. Bloothoof, P. Christen, K. Mandemakers & M. Schraagen (Eds.), *Population Reconstruction* (pp. 61–83). Cham Heidelberg New York Dordrecht London: Springer. doi: [10.1007/978-3-319-19884-2](https://doi.org/10.1007/978-3-319-19884-2)

- Boonstra, O. (1995). Het einde van het analfabetisme. In K. Mandemakers & O. Boonstra (Eds.), *De levensloop van de Utrechtse bevolking in de 19e eeuw* (pp. 68–85). Assen: Van Gorcum.
- Boonstra, O. (2009a). De daling van het analfabetisme. Een regionale analyse over de periode 1775–1900. In O. Boonstra & A. Schuurman (Eds.), *Tijd en ruimte. De toepassing van GIS in de alfawetenschappen* (pp. 62–73). Utrecht: Matrijs.
- Boonstra, O. (2009b). Functioneel analfabetisme in Nederland, 1775–1900. In I. Maas, M. H. D. van Leeuwen & K. Mandemakers (Eds.), *Honderdvijftig jaar levenslopen. De Historische Steekproef Nederlandse bevolking* (pp. 127–148). Amsterdam: Amsterdam University Press.
- Boonstra, O. (2011). "Een echte onderwijzersfamilie". De intergenerationele mobiliteit van onderwijzers en onderwijzeressen, 1811–1922. In Th. Engelen, O. Boonstra & A. Janssens (Eds.), *Levenslopen in transformatie. Liber amicorum bij het afscheid van prof.dr. Paul M.M. Klep* (pp. 13–27). Nijmegen: Valkhof Pers.
- Boonstra, O., & Mandemakers, K. (1995). "Ieder is het kind zijner eigene werken". Sociale stratificatie en mobiliteit in Nederland in de achttiende en negentiende eeuw. In J. Dronkers & W. C. Ultee (Eds.), *Verschuivende ongelijkheid in Nederland. Sociale gelaagdheid en mobiliteit* (pp. 125–141). Assen: Van Gorcum.
- Bosma, U. (2009). European colonial soldiers in the nineteenth century: Their role in white global migration and patterns of colonial settlement. *Journal of Global History*, 4(2), 317–336. doi: [10.1017/S1740022809003179](https://doi.org/10.1017/S1740022809003179)
- Bosma, U. (2010). *Indiëgangsters. Verhalen van Nederlanders die naar Indië trokken*. Amsterdam: Bert Bakker.
- Bosma, U., & Mandemakers, K. (2008). Indiëgangsters: Sociale herkomst en migratiemotieven (1830–1950). Een onderzoek op basis van de Historische Steekproef Nederlandse bevolking (HSN). *Bijdragen en Mededelingen Betreffende de Geschiedenis der Nederlanden*, 123, 162–184. doi: [10.18352/bmgn-lchr.6779](https://doi.org/10.18352/bmgn-lchr.6779)
- Boschaart, M., & Lelieveld, T. (2016). *Historische Steekproef Nederlands-Indië. Een haalbaarheidsonderzoek*. Radboud Universiteit Nijmegen. Retrieved from <https://iisg.amsterdam/en/hsn/products/papers>
- Boter, C. (2017). *Dutch Divergence? Women's work, structural change, and household living standards in the Netherlands, 1830–1914* (PhD thesis Wageningen University). Retrieved from <http://library.wur.nl/WebQuery/wurpubs/fulltext/413345>
- Brandström, A. (2016). Why have historical demographers abandoned history? In K. Matthijs, S. Hin, J. Kok & H. Matsuo (Eds.), *The future of historical demography. Upside down and inside out* (pp. 150–152). Leuven/Den Haag: Acco.
- Bras, H. (1998). Domestic service, migration and the social status of woman at marriage: The case of a Dutch sea province, Zeeland 1820–1935. *Historical Social Research*, 23(3), 3–19. doi: [10.12759/hsr.23.1998.3.3-19](https://doi.org/10.12759/hsr.23.1998.3.3-19)
- Bras, H. (2002). *Zeeuwse meiden. Dieners in de levensloop van vrouwen, ca. 1850–1950* (PhD thesis Utrecht University). Amsterdam: Aksant.
- Bras, H. (2003). Maids to the city: Migration patterns of female domestic servants from the province of Zeeland, the Netherlands (1850–1950). *The History of the Family*, 8(2), 217–246. doi: [10.1016/S1081-602X\(03\)00027-7](https://doi.org/10.1016/S1081-602X(03)00027-7)
- Bras, H. (2004). Social change, the institution of service and youth: the case of service in the lives of rural-born Dutch women, 1840–1940. *Continuity and Change*, 19(2), 241–264. doi: [10.1017/S0268416004004989](https://doi.org/10.1017/S0268416004004989)
- Bras, H. (2011). Intensification of family relations? Changes in the choice of marriage witnesses in the Netherlands, 1830–1950. *TSEG Low Countries Journal of Social and Economic History*, 8(4), 102–135. doi: [10.18352/tseg.347](https://doi.org/10.18352/tseg.347)
- Bras, H. (2014a). Structural and diffusion effects in the Dutch fertility transition, 1870–1940. *Demographic Research*, 30(5) 151–186. doi: [10.4054/DemRes.2014.30.5](https://doi.org/10.4054/DemRes.2014.30.5)
- Bras, H. (2014b). The influence of popular beliefs about childbirth on fertility patterns in mid-twentieth century Netherlands. *Historical Social Research/Historische Sozialforschung*, 39(1), 76–103. doi: [10.12759/hsr.39.2014.1.76-103](https://doi.org/10.12759/hsr.39.2014.1.76-103)
- Bras, H., & Kok, J. (2003). "Naturally, every child was supposed to work". Determinants of the leaving home process in the Netherlands, 1850–1940. In F. van Poppel, M. Oris & J. Lee (Eds.), *The road to independence. Leaving home in Western and Eastern Societies, 16th–20th centuries* (pp. 403–450). Bern: Peter Lang.

- Bras, H., & Kok, J. (2005). "They live in indifference together": Marriage mobility in Zeeland, the Netherlands, 1796–1922. *International Review of Social History*, 50(4), 247–274. doi: [10.1017/S0020859005002130](https://doi.org/10.1017/S0020859005002130)
- Bras, H., & Kok, J. (2016). Diverging pathways? Sibling differences in marriage timing in a commercialized rural region of the Netherlands, 1860–1940. In F. Boudjaaba, C. Dousset & S. Mouysset (Eds.), *Frères et sœurs du Moyen Âge à nos jours* (pp. 189–206). Bern: Peter Lang.
- Bras, H., Kok, J., & Mandemakers, K. (2010). Sibship size and status attainment across contexts: Evidence from the Netherlands, 1840–1925. *Demographic Research*, 23(4), 73–104. doi: [10.4054/DemRes.2010.23.4](https://doi.org/10.4054/DemRes.2010.23.4)
- Bras, H., Liefbroer, A. C., & Elzinga, C. H. (2010). Standardization of pathways to adulthood? An analysis of Dutch cohorts born between 1850 and 1900. *Demography*, 47(4), 1013–1034. doi: [10.1007/BF03213737](https://doi.org/10.1007/BF03213737)
- Bras, H., & Neven, M. (2007). The effects of siblings on the migration of women in two rural areas of Belgium and the Netherlands, 1829–1940. *Population Studies*, 61(1), 53–71. doi: [10.1080/00324720601048319](https://doi.org/10.1080/00324720601048319)
- Bras, H., Van Bavel, J., & Mandemakers, K. (2013). Unraveling the intergenerational transmission of fertility: genetic and shared-environment effects during the demographic transition in the Netherlands, 1810–1910. *The History of the Family*, 18(2), 116–134. doi: [10.1080/1081602X.2013.803491](https://doi.org/10.1080/1081602X.2013.803491)
- Bras, H., van Poppel, F., & Mandemakers, K. (2009). Relatives as spouses: Preferences and opportunities for kin marriage in a Western society. *American Journal of Human Biology*, 21(6), 793–804. doi: [10.1002/ajhb.20896](https://doi.org/10.1002/ajhb.20896)
- Coale, A. J., & Watkins, S. C. (Eds.) (1986). *The decline of fertility in Europe*. Princeton, NJ: Princeton University Press.
- Chotkowski, M. (2006). *Vijftien ladders en een dambord. Contacten van Italiaanse migranten in Nederland, 1860–1940* (Handelseditie dissertatie). Amsterdam: Aksant.
- Courseau, D., Bijak, J., Franck, R., & Silverman, E. (2016). Model-based demography: Towards a research agenda. In A. Grow & J. Van Bavel (Eds.), *Agent-based modelling in population studies. Concepts, methods, and applications. The Springer Series on Demographic Methods and Population Analysis*, 41 (pp. 29–51). Cham: Springer. doi: [10.1007/978-3-319-32283-4\\_2](https://doi.org/10.1007/978-3-319-32283-4_2)
- Delger, H. (2003). Chancen in Rotterdam. Eine Untersuchung der sozialen Mobilität zweier deutscher Migrantengruppen (1870–1879 und 1920–1929). In A. Eder (Ed.), *Wir sind auch da! Über das Leben von und mit Migranten in europäischen Großstädten* (pp. 47–71). Hamburg: Forum Zeitgeschichte 14.
- Delger, H. & Kok, J. (1998). Bridegrooms and biases: A critical look at the study of intergenerational mobility on the basis of marriage certificates. *Historical Methods*, 31(3), 113–121. doi: [10.1080/01615449809601194](https://doi.org/10.1080/01615449809601194)
- Dribe, M., & Scalone, F. (2010). Detecting deliberate fertility control in pre-transitional populations: Evidence from six German villages, 1766–1863. *European Journal of Population*, 26(4), 411–434. doi: [10.1007/s10680-010-9208-8](https://doi.org/10.1007/s10680-010-9208-8)
- Edvinsson, S., & Janssens, A. (2012). Clustering of deaths in families: Infant and child mortality in historical perspective. *Biodemography and Social Biology*, 58(2), 75–86. doi: [10.1080/19485565.2012.738575](https://doi.org/10.1080/19485565.2012.738575)
- Eijkemans, M. J., van Poppel, F., Habbema, D. F., Smith, K. R., Leridon, H., & te Velde, E. R. (2014). Too old to have children? Lessons from natural fertility populations. *Human Reproduction*, 29(6), 1304–1312. doi: [10.1093/humrep/deu056](https://doi.org/10.1093/humrep/deu056)
- Ekamper, P. (2019). *Time and space in historical demography. Some case studies using Dutch micro-data* (PhD thesis University of Groningen). Available from <http://hdl.handle.net/11370/04fdf1a5-03de-488b-ac3a-3c06e6248d50>
- Ekamper, P., Faes, C., & van Poppel, F. (2010). Vrijers die van verre komen, zijn te schromen. Geografische homogamie van huwelijkspartners, 1812–1922. *Mens en Maatschappij*, 85(4), 380–404. doi: [10.5117/MEM2010.4.EKAM](https://doi.org/10.5117/MEM2010.4.EKAM)
- Ekamper, P., van Poppel, F., & Mandemakers, K. (2011). Widening horizons? The geography of the marriage market in nineteenth- and early-twentieth century Netherlands. In M. P. Gutmann, G. D. Deane, E. R. Merchant & K. M. Sylvester (Eds.), *Navigating Time and Space in Population Studies. International Studies in Population*, 9 (pp. 115–160). Dordrecht: Springer. doi: [10.1007/978-94-007-0068-0\\_6](https://doi.org/10.1007/978-94-007-0068-0_6)

- Ekamper, P., van Poppel, F., van Duin, C., & Garssen, J. (2009). 150-years of temperature-related excess mortality in the Netherlands. *Demographic Research*, 21(14), 385–426. doi: [10.4054/DemRes.2009.21.14](https://doi.org/10.4054/DemRes.2009.21.14)
- Ekamper, P., van Poppel, F., van Duin, C., & Mandemakers, K. (2010). Heatwaves and cold spells and their effect on mortality: An analysis of micro-data for the Netherlands in the nineteenth and twentieth centuries. *Annales de Démographie Historique*, 120(2), 55–104. doi: [10.3917/adh.120.0055](https://doi.org/10.3917/adh.120.0055)
- Engelen, Th. (2009). *Van 2 naar 16 miljoen mensen. Demografie van Nederland, 1800–nu*. Amsterdam: Boom. Retrieved from <http://hdl.handle.net/2066/78820>
- Engelen, Th. (2017). What the seasons tell us. The monthly movement of marriages, economic modernization, and secularization in the Netherlands, 1810–1940. *Historical Life Course Studies*, 4, 165–180. Retrieved from <http://hdl.handle.net/10622/23526343-2017-0009?locatt=view:master>
- Engelen, Th. & Hillebrand, H. (1985). Vruchtbaarheid in verandering. Een gezinsreconstructie in Breda, 1850–1940. *Tijdschrift voor Sociale Geschiedenis*, 11(1985), 248–289.
- Engelen, Th. & Kok, J. (2003). Permanent celibacy and late marriage in the Netherlands, 1890–1960. *Population-E*, 58(1), 67–96. doi: [10.3917/popu.301.0069](https://doi.org/10.3917/popu.301.0069)
- Engelen, Th., & van der Woude, A.† (2016). Historical demography in the Netherlands. In A. Fauve-Chamoux, I. Bolovan & S. Sogner (Eds.), *A global history of historical demography. Half a century of interdisciplinarity* (pp. 467–498). Bern: Peter Lang.
- Floud, R., Fogel, R. W., Harris, B., & Hong, S. C. (2011). *The changing body: Health, nutrition, and human development in the western world since 1700*. Cambridge: Cambridge University Press.
- Gerritzen, D. (1998). Voornamen in Zeeland. In K. Mandemakers, O. Hoogerhuis & A. de Klerk (Eds.), *Over Zeeuwse mensen. Demografische en sociale ontwikkelingen in Zeeland in de negentiende en begin twintigste eeuw. Themanummer Zeeland*, 7 (pp. 104–115).
- Gerritzen, D. (2001). Voornamen in Friesland. Over Friese namen, patroniemen en meernamigheid. In J. Frieswijk, L. Jansma, K. Mandemakers & B. Oosterhaven (Eds.), *Oer Fryske minsken/Over Friese mensen Themanummer It Baeken*, 63 (pp. 179–189).
- Giele, J., & van Oenen, G. J. (1974). De sociale structuur van de Nederlandse samenleving rond 1850. *Mededelingenblad. Orgaan van de Nederlandse vereniging tot bevordering van de sociale geschiedenis*, 45, 2–32.
- Gupta, S. (2010). *The study of the impact of early life conditions on later life events. A look across the individual's life course* (PhD thesis Vrije Universiteit Amsterdam). Amsterdam: Thela Thesis. Retrieved from <https://research.vu.nl/en/publications/the-study-of-the-impact-of-early-life-conditions-on-later-life-ev>
- Haarman, S. (2018). *De ontbrekende schakels. Onderzoek naar zoekstrategieën voor cold cases van de Historische Steekproef Nederland (HSN)*. Radboud Universiteit Nijmegen. Retrieved from <https://iisg.amsterdam/en/hsn/products/papers>
- Hajnal, J. (1982). Two kinds of preindustrial household formation system. *Population and Development Review*, 8(3), 449–494. doi: [10.2307/1972376](https://doi.org/10.2307/1972376)
- Hall, P. K., McCaa, R., & Thorvaldsen, G. (Eds.). (2000). *Handbook of international historical microdata for population research*. Minneapolis: Minnesota Population Center.
- Hammel, E. A., & Laslett, P. (1974). Comparing household structure over time and between cultures. *Comparative studies in Society and History*, 16(1), 73–110. doi: [10.1017/S0010417500007362](https://doi.org/10.1017/S0010417500007362)
- Heath, A. (1981). *Social Mobility*. London: Fontana.
- Heerma van Voss, L., & Vermeulen, F. (2000). Living strategies of Twente textile workers, ca. 1920. *HSN Working Paper*, 16.
- Hobcraft, J. (2006). The ABC of demographic behaviour: How the interplay of alleles, brains, and contexts over the life course should shape research aimed at understanding population processes. *Population Studies*, 60(2), 153–187. doi: [10.1080/00324720600646410](https://doi.org/10.1080/00324720600646410)
- Huijsmans, D. P. (2013). *IISG-LINKS dataset historische Nederlandse toponiemen spatio-temporeel 1812–2012, Release 2013.2*.
- Janssens, A. (1998). Class, work and religion in the female life course: The case of a Dutch textile town: Enschede, 1880–1940. *Historical Social Research/Historische Sozialforschung*, 23(1/2), 254–274. doi: [10.12759/hsr.23.1998.1/2.254-274](https://doi.org/10.12759/hsr.23.1998.1/2.254-274)

- Janssens, A. (2009). De rol van vrouwen in de eerste demografische transitie in Nederland. Een vergelijking van twee textielsteden. *TSEG/Low Countries Journal of Social and Economic History*, 6(4), 87–114. doi: [10.18352/tseg.471](https://doi.org/10.18352/tseg.471)
- Janssens, A. (2014). *Labouring lives. Women, work and the demographic transition in the Netherlands, 1880–1960*. Bern: Peter Lang.
- Janssens, A., & Pelzer, B. (2011). De rol van moeders in de overlevingskansen van kinderen in het verleden. Fabrieksmeisjes versus dienstboden, 1880–1940. In Th. Engelen, O. Boonstra & A. Janssens (Eds.), *Levenslopen in transformatie. Liber amicorum bij het afscheid van prof.dr. Paul M.M. Klep* (pp. 76–93). Nijmegen: Valkhof Pers.
- Janssens, A., & Pelzer, B. (2012). Did factory girls make bad mothers? Women's labor market experience, motherhood, and children's mortality risks in the past. *Biodemography and Social Biology*, 58(2), 133–148. doi: [10.1080/19485565.2012.720451](https://doi.org/10.1080/19485565.2012.720451)
- Janssens, A., & Pelzer, B. (2014). Lovely little angels in heaven? The influence of religiously determined cultural life scripts on infant survival in the Netherlands, 1880–1920. *Historical Social Research/ Historische Sozialforschung*, 39(1), 19–47. doi: [10.12759/hsr.39.2014.1.19-47](https://doi.org/10.12759/hsr.39.2014.1.19-47)
- Jennings, J. A., & Gray, C. L. (2014). Climate variability and human migration in the Netherlands, 1865–1937. *Population and Environment*, 36(3), 255–278. doi: [10.1007/s11111-014-0218-z](https://doi.org/10.1007/s11111-014-0218-z)
- Jennings, J. A., & Gray, C. L. (2017). Climate and marriage in the Netherlands, 1871–1937. *Population and Environment*, 38(3), 242–260. doi: [10.1007/s11111-016-0266-7](https://doi.org/10.1007/s11111-016-0266-7)
- Kalmijn, M. (1995). Bruid, bruidegom en bruiloft: Sociale en economische achtergronden van het huwelijkspatroon. In K. Mandemakers & O. Boonstra, *De levensloop van de Utrechtse bevolking in de 19e eeuw* (pp. 86–102). Assen: Van Gorcum.
- Kalmijn, M. (2008). Voorlopers in de echtscheidingsrevolutie. De relatie tussen echtscheiding en sociale klasse in de negentiende en vroeg-twintigste eeuw. In I. Maas, M. H. D. van Leeuwen & K. Mandemakers (Eds.), *Honderdvijftig jaar levenslopen. De Historische Steekproef Nederlandse bevolking* (pp. 81–96). Amsterdam: Amsterdam University Press.
- Kaptijn, R., Thomese, F., Liefbroer, A. C., van Poppel, F., van Bodegom, D., & Westendorp, R. G. J. (2015). The trade-off between female fertility and longevity during the epidemiological transition in the Netherlands. *PLOS One*, 10(12). doi: [10.1371/journal.pone.0144353](https://doi.org/10.1371/journal.pone.0144353)
- Klüsener, S., Scalone, F., & Dribe, M. (2016). Exploring the role of communication in shaping fertility transition patterns in space and time. In A. Grow & J. Van Bavel (Eds.), *Agent-based modelling in population studies. Concepts, methods, and applications* (pp. 369–403). Berlin: Springer.
- Knigge, A. (2015). *Sources of sibling similarity. Status attainment in the Netherlands during modernization* (ICS-Dissertation Utrecht University). Retrieved from <https://dspace.library.uu.nl/handle/1874/312866>
- Knigge, A. (2016). Beyond the parental generation: The influence of grandfathers and great-grandfathers on status attainment. *Demography*, 53(4), 1219–1244. doi: [10.1007/s13524-016-0486-6](https://doi.org/10.1007/s13524-016-0486-6)
- Knigge, A., Maas, I. & van Leeuwen, M. H. D. (2014). Sources of sibling (dis)similarity: Total family impact on status variation in the Netherlands in the nineteenth century. *American Journal of Sociology*, 120(3), 908–948. doi: [10.1086/679104](https://doi.org/10.1086/679104)
- Knigge, A., Maas, I., van Leeuwen, M. H. D., & Mandemakers, K. (2014). Status attainment of siblings during modernization. *American Sociological Review*, 79(3), 549–574. doi: [10.1177/0003122414529586](https://doi.org/10.1177/0003122414529586)
- Knippenberg, H., & de Vos, S. (2008). Vroege ontkerkelijking in Nederland. Een analyse van het geboortecohort 1850–1882. In I. Maas, M. H. D. van Leeuwen & K. Mandemakers (Eds.), *Honderdvijftig jaar levenslopen. De Historische Steekproef Nederlandse bevolking* (pp. 97–126). Amsterdam: Amsterdam University Press.
- Kok, J. (1997). Youth labor migration and its family setting, the Netherlands 1850–1940. *The History of the Family*, 2(4), 507–526. doi: [10.1016/S1081-602X\(97\)90027-0](https://doi.org/10.1016/S1081-602X(97)90027-0)
- Kok, J. (2000). Transitie en trajecten. De levensloopbenadering in de sociale geschiedenis. *Tijdschrift voor Sociale Geschiedenis*, 26(3), 309–329.
- Kok, J. (2004). Choices and constraints in the migration of families: The central Netherlands, 1850–1940. *The History of the Family*, 9(2), 137–158. doi: [10.1016/j.hisfam.2004.01.002](https://doi.org/10.1016/j.hisfam.2004.01.002)

- Kok, J. (2006a). "Eigen baas zijn, da's maar alles". Huwelijksmotivatie van Rotterdammers uit de tweede helft van de negentiende eeuw. In P. van de Laar, L. Lucassen & K. Mandemakers (Eds.), *Naar Rotterdam. Immigratie en levensloop in Rotterdam vanaf het einde van de negentiende eeuw* (pp. 99–118). Amsterdam: Aksant.
- Kok, J. (2006b). Sources for the historical demography of the Netherlands in the 19th and early 20th centuries. In Y.-C. Chuang, Th. Engelen & A. P. Wolf (Eds.), *Positive or preventive. Fertility developments in Taiwan and the Netherlands, 1850–1950* (pp. 41–51). Amsterdam: Aksant.
- Kok, J. (2011). Moeten en mogen trouwen. Gedwongen huwelijken en generatieconflict, 1870–1940. In Th. Engelen, O. Boonstra & A. Janssens (Eds.), *Levenslopen in transformatie. Liber amicorum bij het afscheid van prof.dr. Paul M.M. Klep* (pp. 94–104). Nijmegen: Valkhof Pers.
- Kok, J. (2014a). Historical demography. Understanding temporal change, individual variation and regional persistence. *TSEG/Low Countries Journal of Social and Economic History*, 11(2), 237–260. doi: [10.18352/tseg.140](https://doi.org/10.18352/tseg.140)
- Kok, J. (2014b). "At age 27, she gets furious": Scripts on marriage and life course variation in the Netherlands, 1850–1970. *Historical Social Research/Historische Sozialforschung*, 39(1), 113–132. doi: [10.12759/hsr.39.2014.1.113-132](https://doi.org/10.12759/hsr.39.2014.1.113-132)
- Kok, J. (2017). Church affiliation and life course transitions in the Netherlands, 1850–1970. *Historical Social Research*, 42(2), 59–91. doi: [10.12759/hsr.42.2017.2.59-91](https://doi.org/10.12759/hsr.42.2017.2.59-91)
- Kok, J. (2018). Pigeon pairs and rich folks' dreams. The (lack of) sex preferences for offspring in the Netherlands, 1850–1950. In P. Puschmann & T. Riswick (Eds.), *Building bridges: Scholars, history and historical demography. A festschrift in honor of professor Theo Engelen* (pp. 487–500). Nijmegen: Valkhof Pers.
- Kok, J., Adams, J. W., Ericsson, T., & Moch, L. P. (2002). Autour du livre de Paul-André Rosental. Les sentiers invisibles: espaces, familles et migrations dans la France du XIXe siècle. *Annales de Démographie Historique*, 2002/2(104), 129–144.
- Kok, J., & Bras, H. (2008). Clustering and dispersal of siblings in the North-Holland countryside, 1850–1940. *Historical Social Research/Historische Sozialforschung*, 33(3), 278–300. doi: [10.12759/hsr.33.2008.3.278-300](https://doi.org/10.12759/hsr.33.2008.3.278-300)
- Kok, J., Bras, H., & Roterling, P. (2016). Courtship and bridal pregnancy in the Netherlands, 1870–1950. *Annales de Démographie Historique*, 132(2), 165–191. doi: [10.3917/adh.132.0165](https://doi.org/10.3917/adh.132.0165)
- Kok, J., & Delger, H. (1998). Success or selection? The effect of migration on occupational mobility in a Dutch province, 1840–1950. *Histoire et Mesure*, 13, 289–322. doi: [10.3406/hism.1998.1492](https://doi.org/10.3406/hism.1998.1492)
- Kok, J., Lucassen, L., Kasakoff, A. B., & Schwartz, R. M. (2002). Autour du livre de C. Pooley et J. Turnbull. Migration and mobility in Britain since the XVIIIth century. *Annales de Démographie Historique*, 2002/2(104), 101–123. Retrieved from <https://www.jstor.org/stable/44386102>
- Kok, J., & Mandemakers, K. (2008). Free choice from a limited supply. The marriage market in two Dutch provinces, 1840–1940. *Romanian Journal of Population Studies*, 2(1), 82–104.
- Kok, J., & Mandemakers, K. (2009). "Je zoudt maar last van mij hebben". Verwanten in het Nederlandse huishouden, 1860–1940. *TSEG/Low Countries Journal of Social and Economic History*, 6(4), 139–165. Themanummer 'De Historische Steekproef Nederland (HSN). Van geboortebank tot collaboratory'. doi: [10.18352/tseg.473](https://doi.org/10.18352/tseg.473)
- Kok, J., & Mandemakers, K. (2010). A life-course approach to co-residence in the Netherlands, 1850–1940. *Continuity and Change*, 25(2), 285–312. doi: [10.1017/S0268416010000160](https://doi.org/10.1017/S0268416010000160)
- Kok, J., & Mandemakers, K. (2012). Nuclear hardship in the nuclear heartland? Families and welfare in the Netherlands, 1850–1940. *Workingpaper WOG historische demografie*, WOG/HD/2012-1. Retrieved from [https://soc.kuleuven.be/ceso/historischedemografie/eng/publications\\_working\\_papers\\_wog.html](https://soc.kuleuven.be/ceso/historischedemografie/eng/publications_working_papers_wog.html)
- Kok, J., & Mandemakers, K. (2016). Life and death of singles in cities, 1850–1940. *Journal of Urban History*, 42(1), 101–120. doi: [10.1177/0096144215611098](https://doi.org/10.1177/0096144215611098)
- Kok, J., Mandemakers, K. & Damsma, D. (2010). Atrophied hearts? Partner choice and social reproduction of farmers in the north-western part of the Netherlands in the 19th and early 20th century. In M.-P. Arrizabalaga, I. Bolovan, M. Eppel, J. Kok & M. L. Nagata (Eds.), *Many paths to happiness? Studies in population and family history. A festschrift for Antoinette Fauve-Chamoux* (pp. 362–387). Amsterdam: Aksant.

- Kok, J., Mandemakers, K., & Mönkediek, B. (2014). Flight from the land? Migration flows of the rural population of the Netherlands, 1850–1940. *Espace, Populations, Sociétés* 2014(1), 1–18. Special issue 'Les populations rurales en Europe occidentale du 18ème siècle aux années 1960–1970'. Available from <http://eps.revues.org/5631?lang=en>
- Kok, J., Mandemakers, K., & Wals, H. (2005). City nomads: changing residence as a coping strategy, Amsterdam, 1890–1940. *Social Science History*, 29(1), 15–43. doi: [10.1017/S0145553200013237](https://doi.org/10.1017/S0145553200013237)
- Kok, J., & Van Bavel, J. (2009). Social control and the intergenerational transmission of the age at marriage, rural Holland 1850–1940. *Population-E*, 64(2), 341–360. doi: [10.3917/popu.902.0377](https://doi.org/10.3917/popu.902.0377)
- Kok, J., Vandezande, M., & Mandemakers, K. (2011). Household structure, resource allocation and child well-being. A comparison across family systems. *TSEG/Low Countries Journal of Social and Economic History*, 8(4), 76–101. doi: [10.18352/tseg.346](https://doi.org/10.18352/tseg.346)
- Kok, J., & van Leeuwen, M. H. D. (Eds.) (2005). *Genegenheid en gelegenheid. Twee eeuwen partnerkeuze en huwelijk*. Amsterdam: Aksant.
- Kok, J., Yang, W. S., Hsieh, Y.-H. (2006). 'Marital fertility and birth control in rural Netherlands and Taiwan, 19th and early 20th centuries. In Y.-C. Chuang, T. Engelen & A. P. Wolf (Eds.), *Positive or preventive. Fertility developments in Taiwan and the Netherlands, 1850–1950* (pp. 199–235). Amsterdam: Aksant.
- Kurosu, S. (2016). Historical demography going 'glocal': the Eurasia project and Japan? In K. Matthijs, S. Hin, J. Kok & H. Matsuo (Eds.), *The future of historical demography. Upside down and inside out* (pp. 60–62). Leuven/Den Haag: Acco.
- Lambert, P. S., Zijdeman, R. L., van Leeuwen, M. H. D., Maas, I., & Prandy, K. (2013). The construction of HISCAM: A stratification scale based on social interactions for historical comparative research. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 46(2), 77–89. doi: [10.1080/01615440.2012.715569](https://doi.org/10.1080/01615440.2012.715569)
- Larmuseau, M., Van Geystelen, A., van Oven, M., & Decorte, R. (2013). Genetic genealogy comes of age: Perspectives on the use of deep-rooted pedigrees in human population genetics. *American Journal of Physical Anthropology*, 150(4) 505–511. doi: [10.1002/ajpa.22233](https://doi.org/10.1002/ajpa.22233)
- Lesger, C., Lucassen, L., & Schrover, M. (2002). Is there life outside the migrant network? German immigrants in XIVth century Netherlands and the need for a more balanced migration typology. *Annales de Démographie Historique*, 2002/2(104), 29–50. doi: [10.3917/adh.104.0029](https://doi.org/10.3917/adh.104.0029)
- Lin, X. C. C. C. (2011). *Female heads of households in Eurasian societies. Taipei and Rotterdam in times of industrialization* (PhD thesis Radboud University Nijmegen).
- Lindeboom, M., Portrait, F., & van den Berg, G. (2010). Long-run effects on longevity of a nutritional shock early in life: The Dutch potato famine of 1846–1847. *Journal of Health Economics*, 29(5), 617–629. doi: [10.1016/j.jhealeco.2010.06.001](https://doi.org/10.1016/j.jhealeco.2010.06.001)
- Lucassen, L. (Ed.). (2003). Eindverslag van het NWO-pioniersproject 'De determinanten van het vestigingsproces van immigranten in Nederland, 1860–1960'. Amsterdam: UVA.
- Lucassen, L. (2004). De selectiviteit van blijvers. Een reconstructie van de sociale positie van Duitse migranten in Rotterdam (1870–1885). *TSEG/Low Countries Journal of Social and Economic History*, 1(2), 92–115. doi: [10.18352/tseg.826](https://doi.org/10.18352/tseg.826)
- Lucassen, L. (2005). Huwelijken van Duitse migranten in Nederland (1860–1940). De rol van herkomst, religie, beroep en sekse. *TSEG/Low Countries Journal of Social and Economic History*, 2(2), 54–80. doi: [10.18352/tseg.740](https://doi.org/10.18352/tseg.740)
- Maas, I. (1998). Coding Dutch occupations into ISCO68. The 500 most frequent male occupations in the HSN-Utrecht data set. In: M. H. D. van Leeuwen, I. Maas & A. Miles (Eds.), *Historical international standard coding of occupations: Status quo after coding 500 frequent male occupations* (pp. 95–117). Berlin: Max Planck Institute Berlin.
- Maas, I., & van Leeuwen, M. H. D. (2006). Over dienstboden, landarbeidsters en andere werkende vrouwen. Beroepen van jonge vrouwen en hun moeders in de huwelijksakten van de Zeeuwse Burgerlijke Stand. *Zeeland. Tijdschrift van het Koninklijk Zeeuwsch Genootschap der wetenschappen*, 15(1), 44–59. Special issue K. Mandemakers, L. Hollestelle and A. De Klerk (Eds.), 'Zeeuwen in beweging'.
- Maas, I., & van Leeuwen, M. H. D. (2009). Van een dubbeltje naar een kwartje? Beroepsloopbanen van mannen en vrouwen in Nederland tussen 1865 en 1940. In I. Maas, M. H. D. van Leeuwen & K. Mandemakers (Eds.), *Honderdvijftig jaar levenslopen. De Historische Steekproef Nederlandse bevolking* (pp. 173–202). Amsterdam: Amsterdam University Press.

- Maas, I., & van Leeuwen, M. H. D. (2019). Partner choice in the Netherlands, 1813–1922: The changing importance of ascribed and achieved status. *The History of the Family*, 24(1), 123–148. doi: [10.1080/1081602X.2018.1558412](https://doi.org/10.1080/1081602X.2018.1558412)
- Maas, I., & Zijdeman, R. (2010). Beyond the local marriage market: The influence of modernization on geographical heterogamy. *Demographic Research*, 23(33), 933–962. doi: [10.4054/DemRes.2010.23.33](https://doi.org/10.4054/DemRes.2010.23.33)
- Mandemakers, K. (1989). Historische steekproef Nederlandse bevolking (HSN): Opzet, doelstellingen en internationale context. *Cahier VGI*, 3, 95–107.
- Mandemakers, K. (2000). The Netherlands. Historical Sample of the Netherlands. In P. Kelly Hall, R. McCaa & G. Thorvaldsen (Eds.), *Handbook of international historical microdata for population research* (pp. 149–177). Minneapolis: Minnesota Population Center.
- Mandemakers, K. (2001a). The Historical Sample of the Netherlands (HSN). *Historical Social Research/Historische Sozialforschung*, 26(4), 179–190.
- Mandemakers, K. (2001b). De sociale structuur in Nederland rond 1900. De samenleving in het perspectief van de modernisering 1850–1990. In J. G. S. J. van Maarseveen & P. K. Doorn (Eds.), *Nederland een eeuw geleden geteld. Een terugblik op de samenleving rond 1900* (pp. 185–207). Amsterdam: Stichting Beheer IISG.
- Mandemakers, K. (2004). De Historische Steekproef Nederlandse bevolking (HSN) en het project Life Courses in Context. *Bevolking en Gezin*, 33, 91–114.
- Mandemakers, K. (2006a). Building life course datasets from population registers by the Historical Sample of the Netherlands (HSN). *History and Computing*, 14(1–2), 87–108.
- Mandemakers, K. (2006b). Levensloop in Rotterdam met de Historische Steekproef Nederlandse bevolking. In P. van de Laar, L. Lucassen & K. Mandemakers (Eds.), *Naar Rotterdam. Immigratie en levensloop in Rotterdam vanaf het einde van de negentiende eeuw* (pp. 9–24). Amsterdam: Aksant.
- Mandemakers, K. (2009). *Waarom Jan en Cor met elkaar trouwden: Over grote historische databestanden, koudwatervrees en interdisciplinaire samenwerking*. Amsterdam: Aksant.
- Mandemakers, K. (2017). En hoe zit het eigenlijk met de getuigen? Professionele en relationele achtergrond van de getuigen in de geboorteakten van Nederland, 1812–1922. In Th. Engelen & T. Riswick (Eds.), *Van historische informatiekunde naar historische sensatie. Libretto amicorum voor Onno Boonstra* (pp. 61–78). Nijmegen: Valkhof Pers.
- Mandemakers, K. (forthcoming). The Historical Sample of the Netherlands. Design and Projects. *Historical Life Course Studies*.
- Mandemakers, K., Bloothoofd, G. & Laan, F. (forthcoming). LINKS. The LINKing System for historical family reconstruction in the Netherlands. *Historical Life Course Studies*.
- Mandemakers, K., & Boonstra, O. (Eds.). (1995). *De levensloop van de Utrechtse bevolking in de 19e eeuw*. Assen: Van Gorcum.
- Mandemakers, K., & Dillon, L. (2004). Best practices with large databases on historical populations. *Historical methods*, 37(1), 34–38. doi: [10.3200/HMTS.37.1.34-38](https://doi.org/10.3200/HMTS.37.1.34-38)
- Mandemakers, K., Mönkediek, B., & Kok, J. (2016). De rol van gezin en bedrijf bij de migratie van Nederlandse boerenkinderen tussen 1850 en 1940. In P. Puschmann, R. Paping & K. Matthijs (Eds.), *Familie en levenskansen in het verleden. Jaarboek Historische Demografie 2015* (pp. 25–56). Leuven/Den Haag: Acco.
- Mandemakers, K., Mourits, R. J., Muurling, S., Boter, C., van Dijk, I. K., Maas, I., Van de Putte, B., Zijdeman, R. L., Lambert, P., van Leeuwen, M. H. D., van Poppel, F., & Miles, A. (2018). *HSN standardized, HISCO-coded and classified occupational titles, release 2018.01*.
- Matthijs, K., Hin, S., Kok, J., & Matsuo, H. (Eds.). (2016). *The future of historical demography. Upside down and inside out*. Leuven/Den Haag: Acco. Available from <https://soc.kuleuven.be/ceso/fapos/publications/the-future-of-historical-demography-upside-down-and-inside-out>
- McQuillan, K. (2004). When does religion influence fertility. *Population and Development Review*, 30(1), 25–56. doi: [10.1111/j.1728-4457.2004.00002.x](https://doi.org/10.1111/j.1728-4457.2004.00002.x)
- Meroño-Peñuela, A., Ashkpour, A., van Erp, M., Mandemakers, K., Breure, L., Scharnhorst, A., Schlobach, S., & van Harmelen, F. (2015). Semantic technologies for historical research: A survey. *Semantic web - interoperability, usability, applicability*, 6, 539–564. Retrieved from <http://www.semantic-web-journal.net/content/semantic-technologies-historical-research-survey>

- Mönkediek, B., Kok, J., & Mandemakers, K. (2015). The impact of family setting and local opportunities on leaving home and migration destinations of rural youths, the Netherlands 1860–1940. *Historical Life Course Studies*, 3, 43–65. Retrieved from <http://hdl.handle.net/10622/23526343-2015-0005?locatt=view:master>
- Mourits, R. (2017). Later-life mortality and longevity in late-18th and 19th-century cohorts. Where are we now, and where are we heading? *Historical Life Course Studies*, 4, 1–19. Retrieved from <http://hdl.handle.net/10622/23526343-2017-0001?locatt=view:master>
- Mourits, R. (2019). *Exceptional lives, extraordinary families. Familial clustering of longevity in the 19th and early-20th century* (PhD thesis Radboud University Nijmegen).
- Murphy, M. (2013). The intergenerational transmission of reproductive behaviour: Comparative perspective. *The History of the Family*, 18(2) 107–115. doi: [10.1080/1081602X.2013.808447](https://doi.org/10.1080/1081602X.2013.808447)
- Nomes, E., Grow, A. & Van Bavel, J. (2019). The mid-twentieth century baby boom and the role of social influence. An agent-based modelling approach. *Historical Life Course Studies*, 8, 1–26. Retrieved from: <http://hdl.handle.net/10622/23526343-2019-0001?locatt=view:master>
- Oosten, M. (2008). *Verleden namen* (Master thesis LIACS Leiden). Available from <https://theses.liacs.nl/cs>
- Ory, B., & van Poppel, F. (2013). Trends and risk factors of maternal mortality in late-nineteenth-century Netherlands. *The History of the Family*, 18(4), 481–509. doi: [10.1080/1081602X.2013.836457](https://doi.org/10.1080/1081602X.2013.836457)
- Paiva, D., & Anguita, F. (2017). *Linking the Historical Sample of the Netherlands into the American censuses, 1850–1940*. Abstract retrieved from <http://recordlink.org/abstracts/>
- Popławska, M. (2015). *Essays on insurance and health economics* (PhD thesis Vrije Universiteit Amsterdam). Amsterdam: Tinbergen Institute.
- Puschmann, P. (2015). *Social inclusion and exclusion of urban in-migrants in Northwestern European port cities. Antwerp, Rotterdam & Stockholm ca. 1850–1930* (PhD thesis KU Leuven). Retrieved from [https://limo.libis.be/primo-explore/fulldisplay?docid=LIRIAS1871846&context=L&vid=Lirias&search\\_scope=Lirias&tab=default\\_tab&lang=en\\_US&fromSitemap=1](https://limo.libis.be/primo-explore/fulldisplay?docid=LIRIAS1871846&context=L&vid=Lirias&search_scope=Lirias&tab=default_tab&lang=en_US&fromSitemap=1)
- Puschmann, P., Donrovich, R., Grönberg, P. O., Dekeyser, G., & Matthijs, K. (2016). Disfavored in life, favored in death? Later-life mortality differences (ages 30+) between migrants and natives in Antwerp, Rotterdam and Stockholm, 1850–1930. *Historical Social Research/Historische Sozialforschung*, 41(4), 257–290. doi: [10.12759/hsr.41.2016.4.257-290](https://doi.org/10.12759/hsr.41.2016.4.257-290)
- Puschmann, P., Donrovich, R., & Matthijs, K. (2017). Salmon bias or red herring? Comparing adult mortality risks (ages 30–90) between natives and internal migrants: Stayers, returnees and movers in Rotterdam, the Netherlands, 1850–1940. *Human Nature*, 28(4), 481–499. doi: [10.1007/s12110-017-9303-1](https://doi.org/10.1007/s12110-017-9303-1)
- Puschmann, P., Van den Driessche, N., Grönberg, P. O., Van de Putte, B., & Matthijs, K. (2015). From outsiders to insiders? Partner choice and marriage among internal migrants in Antwerp, Rotterdam & Stockholm, 1850–1930. *Historical Social Research/Historische Sozialforschung*, 4(2), 319–358. doi: [10.12759/hsr.40.2015.2.319-358](https://doi.org/10.12759/hsr.40.2015.2.319-358)
- Quanjer, B., & Kok, J. (2019). Homemakers and heights. Intra-household resource allocation and male stature in the Netherlands, 1860–1930. *Economics and Human Biology*, 34, 192–207. doi: [10.1016/j.ehb.2019.04.003](https://doi.org/10.1016/j.ehb.2019.04.003)
- Quanjer, B., & Kok, J. (2020, forthcoming). Drafting the Dutch, selection biases in Dutch conscript records in the second half of the 19th century. *Social Science History*, 44(4). doi: [10.1017/ssh.2020.13](https://doi.org/10.1017/ssh.2020.13)
- Quaranta, L. (2015). Using the Intermediate Data Structure (IDS) to construct files for statistical analysis. *Historical Life Course Studies*, 2, 86–107. Retrieved from: <http://hdl.handle.net/10622/23526343-2015-0007?locatt=view:master>
- Quaranta, L. (2016). STATA Programs for using the Intermediate Data Structure (IDS) to construct files for Statistical Analysis. *Historical Life Course Studies*, 3, 1–19. Retrieved from: <http://hdl.handle.net/10622/23526343-2016-0001?locatt=view:master>
- Quaranta, L. (2018). Program for studying intergenerational transmissions in infant mortality using the Intermediate Data Structure (IDS). *Historical Life Course Studies*, 7, 11–27. Retrieved from: <http://hdl.handle.net/10622/23526343-2018-0010?locatt=view:master>
- Quaranta, L. & Sommerseth, H. (2018). Introduction: Intergenerational transmissions of infant mortality using the Intermediate Data Structure (IDS). *Historical Life Course Studies*, 7, 1–10. Retrieved from: <http://hdl.handle.net/10622/23526343-2018-00014?locatt=view:master>

- Reher, D. S., Sandström, G., Sanz-Gimeno, A., & van Poppel, F. (2017). Agency in fertility decisions in Western Europe during the Demographic Transition: A comparative perspective. *Demography*, 54(1), 3–22. doi: [10.1007/s13524-016-0536-0](https://doi.org/10.1007/s13524-016-0536-0)
- Riswick, T. (2018). Testing the conditional resource dilution hypothesis: The impact of sibship size and composition on infant and child mortality in the Netherlands, 1863–1910. *The History of the Family*, 23(4), 623–655. doi: [10.1080/1081602X.2018.1532310](https://doi.org/10.1080/1081602X.2018.1532310)
- Riswick, T. (2020). *Between Rivalry and Support: The impact of brothers and sisters on infant and child mortality in the Netherlands (1863–1910) and Taiwan (1906–1945)* (Ph D thesis, Radboud University Nijmegen).
- Rosenbaum-Feldbrügge, M. (2018). Gender differences in response to family crisis: changes in household composition and migration of widowed parents with minor children in the Netherlands, 1863–1910. *The History of the Family*, 23(4), 679–705. doi: [10.1080/1081602X.2018.1537881](https://doi.org/10.1080/1081602X.2018.1537881)
- Rosenbaum-Feldbrügge, M. (2019). *Dealing with demographic stress. Parental death and the transition to adulthood in the Netherlands, 1850–1952* (PhD thesis Radboud University Nijmegen).
- Rosenbaum-Feldbrügge, M., & Debiasi, E. (2019). The impact of parental death on the timing of first marriage: Evolutionary versus social explanations (the Netherlands, 1850–1940). *Demographic Research*, 40(28), 799–834. doi: [10.4054/DemRes.2019.40.28](https://doi.org/10.4054/DemRes.2019.40.28)
- Rotering, P., & Bras, H. (2015). With the help of kin? Household composition and reproduction in the Netherlands 1842–1920. *Human Nature*, 26(1), 102–121. doi: [10.1007/s12110-015-9222-y](https://doi.org/10.1007/s12110-015-9222-y)
- Schalk, R. (2015). *Splitting the bill: matching schooling to Dutch labour markets, 1750–1920* (PhD thesis Utrecht University). Retrieved from <https://dspace.library.uu.nl/handle/1874/345716>
- Schellekens, J., & van Poppel, F. (2006). Religious differentials in marital fertility in The Hague (Netherlands) 1860–1909. *Population Studies*, 60(1), 23–28. doi: [10.1080/00324720500430758](https://doi.org/10.1080/00324720500430758)
- Schellekens, J., & van Poppel, F. (2012). Marital fertility decline in the Netherlands: Child mortality, real wages and unemployment, 1860–1939. *Demography*, 49(3), 965–988. doi: [10.1007/s13524-012-0112-1](https://doi.org/10.1007/s13524-012-0112-1)
- Schellekens, J., & van Poppel, F. (2016). Early-life conditions and adult mortality decline in Dutch cohorts born 1812–1921. *Population Studies*, 70(3), 327–343. doi: [10.1080/00324728.2016.1223336](https://doi.org/10.1080/00324728.2016.1223336)
- Schenk, N., & van Poppel, F. (2011). Social class, social mobility and mortality in the Netherlands, 1850–2004. *Explorations in Economic History*, 48(3), 401–417. doi: [10.1016/j.eeh.2011.05.010](https://doi.org/10.1016/j.eeh.2011.05.010)
- Schraagen, M. (2014). *Aspects of Record Linkage* (PhD thesis Leiden University). Retrieved from <http://hdl.handle.net/1887/29716>
- Schraagen, M., & Hoogeboom, H. J. (2011). Predicting record linkage potential in a family reconstruction graph. In P. De Causmaecker, J. Maervoet, T. Messelis, K. Verbeeck & T. Vermeulen (Eds.), *Proceedings of 23rd Benelux Conference on Artificial Intelligence (BNAIC 2011)* (pp. 199–206).
- Schraagen, M., & Huijsmans, N. (2013). Comparison between historical population archives and decentralized databases. In *Proceedings of the 7th workshop on language technology for cultural heritage, social sciences, and humanities (LaTeCH) (2013)* (pp. 20–28). Retrieved from <https://www.aclweb.org/anthology/W13-2703>
- Schraagen, M., & Kusters, W. (2014). Record linkage using graph consistency. In P. Perner (Ed.), *Machine learning and data mining in pattern recognition. MLDM 2014. Lecture notes in computer science*, 8556 (pp. 471–483). Cham: Springer. doi: [10.1007/978-3-319-08979-9\\_36](https://doi.org/10.1007/978-3-319-08979-9_36)
- Schrover, M. (2000a). Potverkopers, vijlenkappers, winkeliers en stukadoors. Nichevorming onder Duitse migranten in de negentiende-eeuwse stad Utrecht. *Tijdschrift voor Sociale Geschiedenis*, 26(4), 281–305.
- Schrover, M. (2000b). German communities in nineteenth-century Utrecht: Factors influencing the settlement process. *IMIS-Beitrage*, 14, 45–74. Retrieved from <https://openaccess.leidenuniv.nl/handle/1887/13923>
- Schrover, M. (2001). Immigrant business and niche formation in a historical perspective. The Netherlands in the nineteenth century. *Journal of Ethnic and Migration Studies*, 27, 295–311. Retrieved from <https://openaccess.leidenuniv.nl/handle/1887/13905>
- Schrover, M. (2002). *Een kolonie van Duitsers. Groepsvorming onder Duitse immigranten in Utrecht in de negentiende eeuw*. Amsterdam: Aksant.
- Schrover, M. (2003). Living together, working together: Concentrations amongst German immigrants in the Netherlands in the nineteenth century. *Continuity and Change*, 18(2), 263–285. doi: [10.1017/S0268416003004570](https://doi.org/10.1017/S0268416003004570)
- Schrover, M. (2004). "De Duitse is immers tegenwoordig ook zeer galant voor Dames". Huwelijksgedrag van migranten in Nederland in de negentiende eeuw. *Bevolking en Gezin* 33(2), 103–125.

- Schrover, M. (2005). Huwelijk, gender, migratie en integratie. Partnerkeuze van Duitsers in Utrecht in de negentiende eeuw. In J. Kok & M. H. D. van Leeuwen (Eds.), *Genegenheid en gelegenheid. Twee eeuwen partnerkeuze en huwelijk* (pp. 135–158). Amsterdam: Aksant. Retrieved from [https://www.ru.nl/publish/pages/800120/geng\\_totaal.pdf](https://www.ru.nl/publish/pages/800120/geng_totaal.pdf)
- Schulz, W. (2013). *Careers of men and women in the 19th and 20th centuries* (PhD thesis University Utrecht). ICS Dissertation Series 207. Retrieved from <https://dspace.library.uu.nl/handle/1874/279589>
- Schulz, W. (2015). Occupational career attainment of single women during modernization: The logic of industrialism thesis revisited. *European Societies*, 7(4), 467–491. doi: [10.1080/14616696.2015.1072228](https://doi.org/10.1080/14616696.2015.1072228)
- Schulz, W., & Maas, I. (2010). Studying historical occupational careers with multilevel growth models. *Demographic Research*, 23(24), 669–696. doi: [10.4054/DemRes.2010.23.24](https://doi.org/10.4054/DemRes.2010.23.24)
- Schulz, W., & Maas, I. (2012). Studying career success: The role of resources and norms for occupational status attainment in the Netherlands, 1865–1940. *European Sociological Review*, 28(2), 220–240. doi: [10.1093/esr/jcq060](https://doi.org/10.1093/esr/jcq060)
- Schulz, W., Maas, I., & van Leeuwen, M. H. D. (2014). When women disappear from the labour market: Occupational status of Dutch women at marriage in a modernizing society, 1865–1922. *The History of the Family*, 19(4), 426–446. doi: [10.1080/1081602X.2014.955516](https://doi.org/10.1080/1081602X.2014.955516)
- Schulz, W., Maas, I., & van Leeuwen, M. H. D. (2015). Occupational career attainment during modernization. A study of Dutch men in 841 municipalities between 1865 and 1928. *Acta Sociologica*, 58(1), 5–24. doi: [10.1177/0001699314565795](https://doi.org/10.1177/0001699314565795)
- Sear, R., & Coall, D. (2011). How much does family matter? Cooperative breeding and the demographic transition. *Population and Development Review*, 37, 81–112. doi: [10.1111/j.1728-4457.2011.00379.x](https://doi.org/10.1111/j.1728-4457.2011.00379.x)
- Shepherd, J.R., Kok, J., & Hsieh, Y. (2006). Fertility and infant and early childhood mortality: Some lessons on stopping behaviour from Taiwanese and Dutch cases. In Y.-C. Chuang, T. Engelen & A. P. Wolf (Eds.), *Positive or preventive. Fertility developments in Taiwan and the Netherlands, 1850–1950* (pp. 163–197). Amsterdam: Aksant.
- Shepherd, J.R., Pan, I., Kok, J., Engel, C., Engelen, T., & Brown, M. (2006). Group identity and fertility: An evaluation of the role of religion and ethnicity in the Netherlands and Taiwan. In Y.-C. Chuang, T. Engelen & A. P. Wolf (Eds.), *Positive or preventive. Fertility developments in Taiwan and the Netherlands, 1850–1950* (pp. 121–161). Amsterdam: Aksant.
- Shepherd, J. R., Schoonheim, M., Chang, T., & Kok, J. (2011). Maternal mortality in Taiwan and the Netherlands, 1850–1945. In T. Engelen, J. R. Shepherd & W.-S. Yang (Eds.), *Death at the opposite ends of the Eurasian continent. Mortality trends in Taiwan and the Netherlands 1850–1945* (pp. 229–273). Amsterdam: Aksant.
- Smith, K. R., Hanson, H. A., & Mineau, G. P. (2016). A six-pack of big ideas for historical demography. In K. Matthijs, S. Hin, J. Kok & H. Matsuo (Eds.), *The future of historical demography. Upside down and inside out* (pp. 257–260). Leuven/Den Haag: Acco.
- Smits, L. J. (1998). *Preconceptional determinants of the fecundity of female offspring* (PhD thesis KU Nijmegen). Available from <http://hdl.handle.net/2066/146649>
- Smits, L. J., Jongbloet, P. H., & Zielhuis, G. A. (2000). Fecundity of daughters born after short, intermediate, or long birth intervals. An analysis of family reconstitutions from the Netherlands, 19th and early 20th century. *Social Biology*, 47(1–2), 18–33. doi: [10.1080/19485565.2000.9989007](https://doi.org/10.1080/19485565.2000.9989007)
- Smits, L. J., Jongbloet, P. H., & Zielhuis, G. A. (2001). Season of birth and reproductive performance. An analysis of family reconstitutions of 800 women born at the end of the 19th century. *Chronobiology International*, 18(3), 525–539. doi: [10.1081/CBI-100103974](https://doi.org/10.1081/CBI-100103974)
- Smits, L. J., Zielhuis, G. A., Jongbloet, P. H., & Straatman, H. (1998). Seasonal variation in human fecundability. *Human Reproduction*, 13(12), 3520–3524. doi: [10.1093/humrep/13.12.3520](https://doi.org/10.1093/humrep/13.12.3520)
- Smits, L. J., Zielhuis, G. A., Jongbloet, P. H., & van Poppel, F. W. A. (2002). Mother's age and daughter's fecundity. An epidemiological analysis of late 19th to early 20th century family reconstitution. *International Journal of Epidemiology*, 31(2), 349–358. Retrieved from: <https://www.ncbi.nlm.nih.gov/pubmed/11980796>
- Sprok, J. J. A. (2013). Status attainment in the Netherlands during the period 1851–1890: The influence of social capital on status attainment in the context of modernization. *Social Cosmos*, 4(2), 152–166. Available from <https://dspace.library.uu.nl/handle/1874/288644>

- Suanet, B., & Bras, H. (2014). Sibling position and marriage timing in the Netherlands, 1840–1922: A comparison across social classes, local contexts, and time. *Journal of Family History*, 39(2), 126–139. doi: [10.1177/0363199013506986](https://doi.org/10.1177/0363199013506986)
- Tammes, P. (2010). Jewish-Gentile intermarriage in pre-war Amsterdam. *The History of the Family*, 15(3), 298–315. doi: [10.1016/j.hisfam.2010.05.003](https://doi.org/10.1016/j.hisfam.2010.05.003)
- Tammes, P. (2011). Residential segregation of Jews in Amsterdam on the eve of the Shoah. *Continuity and Change*, 26(2), 243–270. doi: [10.1017/S0268416011000129](https://doi.org/10.1017/S0268416011000129)
- Tammes, P. (2012a). “Hack, pack, sack”: Occupational structure, status and mobility of Jews in Amsterdam, 1851–1941. *Journal of Interdisciplinary History*, 43(1), 1–26. doi: [10.1162/JINH\\_a\\_00336](https://doi.org/10.1162/JINH_a_00336)
- Tammes, P. (2012b). Abandoning Judaism: A life history perspective on disaffiliation and conversion to Christianity among prewar Amsterdam Jews. *Advances in Life Course Research*, 17(2), 81–92. doi: [10.1016/j.alcr.2012.02.001](https://doi.org/10.1016/j.alcr.2012.02.001)
- Tammes, P., & Scholten, P. (2017). Assimilation of ethnic-religious minorities in the Netherlands: A historical-sociological analysis of pre-World War II Jews and contemporary Muslims. *Social Science History*, 41(3), 477–504. doi: [10.1017/ssh.2017.12](https://doi.org/10.1017/ssh.2017.12)
- Tammes, P., & van Poppel, F. (2012). The impact of assimilation on the family structure of Jews in Amsterdam, 1880–1940. *Journal of Family History*, 37(4), 395–416. doi: [10.1177/0363199012442470](https://doi.org/10.1177/0363199012442470)
- Tassenaar, V. (2019). Development of regional variety of the biological standard of living in the Netherlands, 1812–1913. *Economics & Human Biology*, 34, 151–161. doi: [10.1016/j.ehb.2019.02.004](https://doi.org/10.1016/j.ehb.2019.02.004)
- Thernstrom, S. (1973). *The other Bostonians. Poverty and progress in the American metropolis, 1880–1970*. Cambridge: Harvard University Press.
- Thompson, K., Lindeboom, M., & Portrait, F. (2019). Adult body height as a mediator between early-life conditions and socio-economic status: The case of the Dutch Potato Famine, 1846–1847. *Economics & Human Biology*, 34, 103–114. doi: [10.1016/j.ehb.2019.04.006](https://doi.org/10.1016/j.ehb.2019.04.006)
- Van Bavel, J. (2001). *Demografische reproductie en sociale evolutie: Geboortebeperving in Leuven 1846–1910* (PhD thesis Katholieke Universiteit Leuven).
- Van Bavel, J., & Kok, J. (2004). Birth spacing in the Netherlands. The effects of family composition, occupation and religion on birth intervals, 1820–1885. *European Journal of Population*, 20(2), 119–140. doi: [10.1023/B:EUJP.0000033860.39537.e2](https://doi.org/10.1023/B:EUJP.0000033860.39537.e2)
- Van Bavel, J., & Kok, J. (2005a). The role of religion in the Dutch fertility transition: Starting, spacing, and stopping in the heart of the Netherlands, 1845–1945. *Continuity and Change*, 20(2), 247–263. doi: [10.1017/S0268416005005473](https://doi.org/10.1017/S0268416005005473)
- Van Bavel, J., & Kok, J. (2005b). Are deliberate birth spacing effects in fact statistical and bio-demographic artifacts? A critical study with data from the Netherlands. *Unpublished paper XXV IUSSP International Population Conference, Tours, July 2005*.
- Van Bavel, J., & Kok, J. (2010a). Pioneers of the modern lifestyle? Childless couples in the early twentieth-century Netherlands. *Social Science History*, 34(1), 47–72. Retrieved from <https://www.jstor.org/stable/40587335>
- Van Bavel, J., & Kok, J. (2010b). Uitstel en afstel van ouderschap tijdens het interbellum in Nederland. In J. Kok & J. Van Bavel (Eds.), *De levenskracht der bevolking. Sociale en demografische kwesties in de Lage Landen tijdens het interbellum* (pp. 197–227). Leuven: Universitaire Pers Leuven.
- Van Bavel, J., & Kok, J. (2010c). A mixed effects model of birth spacing for pre-transition populations. Evidence of deliberate fertility control from nineteenth century Netherlands. *The history of the Family*, 15(2), 125–138. doi: [10.1016/j.hisfam.2009.12.004](https://doi.org/10.1016/j.hisfam.2009.12.004)
- Van Bavel, J., Kok, J., & Engelen, Th. (2008). Hoge kinderloosheid tijdens het interbellum in Nederland. De rol van godsdienst, levensstandaard en economische crisis. In I. Maas, M. H. D. van Leeuwen & K. Mandemakers (Eds.), *Honderdvijftig jaar levenslopen. De Historische Steekproef Nederlandse bevolking* (pp. 51–80). Amsterdam: Amsterdam University Press.
- van Bruggen, S., Bulten, L., & Maas, M. (2017). *Pilotonderzoek Historische Steekproef Nederlands-Indië. Eindrapport projectstage Big Data*. Radboud Universiteit Nijmegen. Retrieved from <https://iisg.amsterdam/en/hsn/products/papers>
- van de Laar, P., Lucassen, L., & Mandemakers, K. (Eds.). (2006). *Naar Rotterdam. Immigratie en levensloop in Rotterdam vanaf het einde van de negentiende eeuw*. Amsterdam: Aksant.
- Van de Putte, B., & Miles, A. (2005). A social classification scheme for historical occupational data. *Historical Methods*, 38(2), 61–94. doi: [10.3200/HMTS.38.2.61-94](https://doi.org/10.3200/HMTS.38.2.61-94)

- van den Belt, A., Kok, J., & Mandemakers, K. (2011). Digital thombos: A new source for 18th century Sri Lankan family history. Research note. *The History of the Family*, 16(4), 481–489. doi: [10.1016/j.hisfam.2011.08.006](https://doi.org/10.1016/j.hisfam.2011.08.006)
- van den Berg, G. & Gupta, S. (2015). The role of marriage in the causal pathway from economic conditions early in life to mortality. *Journal of Health Economics*, 40, 141–158. doi: [10.1016/j.jhealeco.2014.02.004](https://doi.org/10.1016/j.jhealeco.2014.02.004)
- van den Berg, G., Lindeboom, M., & Lopez, M. (2006). Economic conditions early in life and individual mortality. *American Economic Review*, 96(1), 290–302. doi: [10.1257/000282806776157740](https://doi.org/10.1257/000282806776157740)
- van den Berg, G., Lindeboom, M., & Portrait, F. (2013). The Dutch Potato Famine 1846–1847: A study on the relationships between early-life exposure and later-life mortality. In L. H. Lumey & A. Vaiserman (Eds.), *Early life nutrition and adult health and development: Lessons from changing dietary patterns, famines, and experimental studies* (pp. 229–250). New York: Nova Science Publishers.
- van den Berg, N. (2020). *Family matters: a genealogical inquiry into the familial component of longevity* (PhD thesis University of Leiden).
- van den Berg, N., Rodriguez-Girondo, M., Mandemakers, K., Janssens, A. A .P. O., Beekman, M., & Slagboom, P. E. (2019a). Longevity relatives count score defines heritable longevity carriers and suggest case improvement in genetic studies. *BioRxiv*. doi: [10.1101/609891](https://doi.org/10.1101/609891)
- van den Berg, N., Rodriguez-Girondo, M., van Dijk, I. K., Mourits, R. J., Mandemakers, K., Janssens, A. A .P. O., Beekman, M., Smith, K. R., & Slagboom, P. E. (2019b). Longevity defined as 10% survivors and beyond is transmitted as a quantitative genetic trait. *Nature Communications*, 10(35), 1–12. doi: [10.1038/s41467-018-07925-0](https://doi.org/10.1038/s41467-018-07925-0)
- van den Berg, N., van Dijk, I. K., Mourits, R. J., Slagboom, P. E., Janssens, A. P. O., & Mandemakers, K. (2020). Families in comparison: An individual-level comparison of life-course and family reconstructions between population and vital event registers. *Population Studies*. doi: [10.1080/00324728.2020.1718186](https://doi.org/10.1080/00324728.2020.1718186)
- van der Harst, G. (2006). Brabanders en Zeeuwen in het Rotterdamse straatbeeld. De verhuisgeschiedenis van de eerste generatie immigranten, 1870–1920. In P. van de Laar, L. Lucassen & K. Mandemakers (Eds.), *Naar Rotterdam. Immigratie en levensloop in Rotterdam vanaf het einde van de negentiende eeuw* (pp. 39–60). Amsterdam: Aksant.
- van der Meer, A., & Boonstra, O. (2006). *Repertorium van Nederlandse gemeenten 1812–2006*. Den Haag: DANS.
- van der Velden, S. (2012). A unique branch of the working class? Dutch seaman 1900–1940. In R. Gorski & B. Söderqvist (Eds.), *The parallel worlds of the seafarer: Ashore, afloat and abroad. Papers from the 10th North Sea History Conference* (pp. 119–139). Göteborg: Sjöfartsmuseet Akvariet.
- van der Woude, A. M. (1980). Socioculturele geschiedenis 1500–1800. De alfabetisering. In *Algemene Geschiedenis der Nederlanden, VII* (pp. 256–264). Haarlem: Fibula-Van Dishoeck.
- Vandezande, M., Kok, J., & Mandemakers, K. (2011). The impact of kin co-residence on survival chances of infants and children in the Netherlands, 1863 –1909. A new methodological approach. *Romanian Journal of Population Studies*, 5(2), 139–160.
- Vandezande, M., Matthijs, K., & Kok, J. (2011). Sibling struggle for schooling: Between resources dilution and collaborative learning, the Netherlands, 1850–1920. *Historical Social Research/ Historische Sozialforschung*, 36(3), 330–349. Retrieved from <https://www.jstor.org/stable/23032439>
- van Dijk, I. K. (2019). *Death and the family. High mortality families and the life course, the Netherlands, 1812–1912* (PhD thesis Radboud Universiteit Nijmegen).
- van Dijk, I. K., Janssens, A., & Smith, K. R. (2018). The long harm of childhood? Life expectancy of adult survivors of high mortality families. *European Journal of Population*, 2018, 1–21. doi: [10.1007/s10680-018-9505-1](https://doi.org/10.1007/s10680-018-9505-1)
- van Dijk, I. K., & Mandemakers, K. (2018). Like mother, like daughter. Intergenerational transmission of infant mortality clustering in Zeeland, the Netherlands, 1833–1912. *Historical Life Course Studies*, 7, 28–46. Retrieved from <http://hdl.handle.net/10622/23526343-2018-0003?locatt=view:master>
- van Gaalen, R. I. (2007). *Solidarity and ambivalence in parent-child relationships* (PhD thesis Utrecht University). Retrieved from <https://www.semanticscholar.org/paper/Solidarity-and-ambivalence-in-parent-child-Gaalen/31b58ed23ba30a7b3d4ec5b9ae7cad2692d1922d>

- van Gaalen, R., & van Poppel, F. (2009). Long-term changes in the living arrangements of children in the Netherlands. *Journal of Family Issues*, 30(5), 653–669. doi: [10.1177/0192513X08331116](https://doi.org/10.1177/0192513X08331116)
- van Galen, C. W. & Hassankhan, M. S. (2018). A research-note on the slave registers of Suriname, 1830–1865. *The History of the Family*, 23 (3), 503–520. doi: [10.1080/1081602X.2018.1507917](https://doi.org/10.1080/1081602X.2018.1507917)
- van Leeuwen, M. H. D., & Maas, I. (1995). Groeiende openheid van de Nederlandse samenleving: Een nieuw fenomeen of een lange trend? Intergenerationele huwelijks- en carrièremobiliteit in de provincie Utrecht, 1850–1940. *Mens en Maatschappij*, 70(4), 321–333. Retrieved from <https://ugp.rug.nl/MenM/article/view/18445/0>
- van Leeuwen, M. H. D., & Maas, I. (1997). Social mobility in a Dutch province, Utrecht 1850–1940. *Journal of Social History*, 30(3), 619–644. doi: [10.1353/jsh/30.3.619](https://doi.org/10.1353/jsh/30.3.619)
- van Leeuwen, M. H. D., & Maas, I. (2005). Endogamy and social class in history: An overview. *International Review of Social History*, 50(S13), 1–23. doi: [10.1017/S002085900500204X](https://doi.org/10.1017/S002085900500204X)
- van Leeuwen, M. H. D., & Maas, I. (2007a). Economische specialisering en veranderende sociale verhoudingen in de 19e en 20e eeuw. Een studie op basis van de Nederlandse volkstellingen en huwelijksakten. In O. W. A. Boonstra, P. K. Doorn, M. P. M. van Horik, J. G. S. J. van Maarseveen & J. Oudhof (Eds.), *Twee eeuwen Nederland geteld. Onderzoek met de digitale Volks- Beroeps- en Woningtellingen 1795–2001* (pp. 181–205). Den Haag: DANS Symposium publicaties 2. Retrieved from <https://www.knaw.nl/shared/resources/actueel/publicaties/pdf/twee-eeuwen-nederland-geteld-Onderzoek-met-de-digitale-volks-beroeps-en-Woningtellingen-1795-2001/view>
- van Leeuwen, M. H. D., & Maas, I. (2007b). Herhaald huwen: Een historische studie naar heterogamie. In T. van der Lippe, P. A. Dykstra, G. Kraaykamp & J. Schippers (Eds.), *De maakbaarheid van de levensloop* (pp. 43–58). Assen: Van Gorcum.
- van Leeuwen, M. H. D., & Maas, I. (2011). *HISCLASS: A Historical Social Class scheme*. Leuven: Leuven University Press.
- van Leeuwen, M. H. D., & Maas, I., & Mandemakers, K. (2005). Het kiezen van een huwelijkspartner in Nederland 1840–1940: De rol van de familie. In J. Kok & M. H. D. van Leeuwen (Eds.), *Genegenheid en gelegenheid. Twee eeuwen partnerkeuze en huwelijk* (pp. 63–84). Amsterdam: Aksant.
- van Leeuwen, M. H. D., & Maas, I., Miles, A. (2002). *HISCO. Historical International Standard Classification of Occupations*. Leuven: Leuven University Press.
- van Leeuwen, M. H. D., & Maas, I., Miles, A. (2004). Creating a historical standard classification of occupations. An exercise in multinational, interdisciplinary cooperation. *Historical Methods*, 37(4), 186–197. doi: [10.3200/HMTS.37.4.186-197](https://doi.org/10.3200/HMTS.37.4.186-197)
- van Leeuwen, M. H. D., & Zijdeman, R. (2010). Digital humanities and the history of working women: A cascade. *The History of the Family*, 19(4). doi: [10.1080/1081602X.2014.955517](https://doi.org/10.1080/1081602X.2014.955517)
- van Maarseveen, J. G. S. J. (2002). Beroepstellingen 1849–1971/2000. In J. G. S. J. van Maarseveen (Ed.), *Algemene tellingen in de twintigste eeuw* (pp. 115–144). Voorburg/Heerlen: Centraal Bureau voor de Statistiek.
- van Poppel, F. (1995). Overleven in Utrecht. Levensduur en sociale ongelijkheid in negentiende-eeuwse geboortegeneraties. In K. Mandemakers & O. Boonstra (Eds.), *De levensloop van de Utrechtse bevolking in de 19e eeuw* (pp. 37–67). Assen: Van Gorcum.
- van Poppel, F., Bijwaard, G. E., Ekamper, P., & Mandemakers, K. (2012). Historical trends in the correlation of sibling deaths in infancy in the Netherlands. *Biodemography and Social Biology*, 58(2), 87–115. doi: [10.1080/19485565.2012.720448](https://doi.org/10.1080/19485565.2012.720448)
- van Poppel, F., Ekamper, P., & Mandemakers, K. (2018). Season of birth and early childhood mortality: A review of the debate and a case study for the Netherlands, 1812–1912. In P. Puschmann & T. Riswick (Eds.), *Building bridges. Scholars, history and historical demography. A festschrift in honor of professor Theo Engelen* (pp. 590–625). Nijmegen: Valkhof Pers.
- van Poppel, F., Ekamper, P., & van Solinge, H. (2007). Farmer looking for a wife. Marital behavior of the farming population in nineteenth-century Netherlands. In H. Moerbeek, A. Niehof & J. van Ophem (Eds.), *Changing families and their lifestyles* (pp. 51–76). Wageningen: Wageningen Academic Publishers.
- van Poppel, F., Jennissen, R., & Mandemakers, K. (2009). Time trends in social class mortality differentials in the Netherlands 1820–1920. An assessment based on indirect estimation techniques. *Social Science History*, 33(2), 119–153. doi: [10.1017/S0145553200010932](https://doi.org/10.1017/S0145553200010932)
- van Poppel, F., Jonker, M., & Mandemakers, K. (2005). Differential infant and child mortality in three Dutch provinces, 1812–1909. *Economic History Review*, 58(2), 272–309. doi: [10.1111/j.1468-0289.2005.00305.x](https://doi.org/10.1111/j.1468-0289.2005.00305.x)

- van Poppel, F., Kok, J., & Kruse, E. (1997). Mortality among illegitimate children in mid-nineteenth-century The Hague. In C. A. Corsini & P. P. Viazzo (Eds.), *The decline of infant and child mortality. The European experience: 1750–1990* (pp. 193–211). The Hague: Martinus Nijhoff Publishers.
- van Poppel, F., Liefbroer, A. C., & Schellekens, J. (2003). Religion and social mobility in nineteenth-century The Hague. *Sociology of Religion*, 64(2), 247–272. doi: [10.2307/3712373](https://doi.org/10.2307/3712373)
- van Poppel, F., & Mandemakers, K. (1997). Differential infant and child mortality in the Netherlands 1812–1912: First results of the historical sample of the population of the Netherlands. In A. Bideau, B. Desjardins & H. Pirez-Brignoli (Eds.), *Infant and child mortality in the past* (pp. 276–300). Oxford: Clarendon Press.
- van Poppel, F., & Mandemakers, K. (2002). Sociale verschillen in zuigelingen- en kindersterfte in Nederland: 1812-1912. *Bevolking en Gezin* 31(2), 5–39.
- van Poppel, F., Monden, C., & Mandemakers, K. (2008). Marriage timing over the generations. *Human Nature*, 19(1), 7–22. doi: [10.1007/s12110-008-9033-5](https://doi.org/10.1007/s12110-008-9033-5)
- van Poppel, F., Reher, D., Sanz-Gimeno, A., Sanchez Dominguez, M., & Beekink, E. (2012). Mortality decline and reproductive change during the Dutch demographic transition: Revisiting a traditional debate with new data. *Demographic Research*, 27(11), 299–338. doi: [10.4054/DemRes.2012.27.11](https://doi.org/10.4054/DemRes.2012.27.11)
- van Poppel, F., Schellekens, J., & Liefbroer, A. C. (2002). Religious differentials in infant and child mortality in Holland. *Population Studies*, 56(3), 277–289. doi: [10.1080/00324720215932](https://doi.org/10.1080/00324720215932)
- van Poppel, F., Schellekens, J., & Walhout, E. (2009). Oversterfte van jonge meisjes in Nederland in de negentiende en eerste helft twintigste eeuw. *TSEG/Low Countries Journal of Social and Economic History*, 6(4), 37–69. doi: [10.18352/tseg.469](https://doi.org/10.18352/tseg.469)
- van Poppel, F., Schenk, N., & van Gaalen, R. (2013). Demographic transitions and changes in the living arrangements of children: The Netherlands 1850–2010. *Population Research and Policy Review*, 32(2), 243–260. doi: [10.1007/s11113-012-9264-3](https://doi.org/10.1007/s11113-012-9264-3)
- van Poppel, F., & Schoonheim, M. (2005). Measuring cultural differences between religions using network data. An example based on nineteenth-century Dutch marriage certificates. *Annales de Démographie Historique*, 2005/1(109), 173–197. doi: [10.3917/adh.109.0173](https://doi.org/10.3917/adh.109.0173)
- van Poppel, F., Tammes, P., & Schenk, N. (2010). Opgroeien in de stad en op het platteland: Verschillen in de gezinssituatie van kinderen in de late negentiende en vroegetwintigste eeuw. In K. Matthijs, B. Van de Putte, J. Kok & H. Bras (Eds.), *Leven in de Lage Landen: Historisch-demografisch onderzoek in Vlaanderen en Nederland* (pp. 53–71). Leuven/Den Haag: Acco.
- van Poppel, F., van Dalen, H. P., & Walhout, E. (2009). Diffusion of a social norm: Tracing the emergence of the housewife in the Netherlands, 1812–1922. *Economic History Review*, 62, 99–127. doi: [10.1111/j.1468-0289.2008.00433.x](https://doi.org/10.1111/j.1468-0289.2008.00433.x)
- van Poppel, F., & van Gaalen, R. (2008a). [Sociale klasse, sociale mobiliteit en sterfte in Nederland, 1850–2007](#). In I. Maas, M. H. D. van Leeuwen & K. Mandemakers (Eds.), *Honderdvijftig jaar levenslopen. De Historische Steekproef Nederlandse bevolking* (pp. 203–236). Amsterdam: Amsterdam University Press.
- van Poppel, F., & van Gaalen, R. (2008b). The presence of parents and childhood survival: The passage of social time and differences by social class. In T. Bengtsson & G. P. Mineau (Eds.), *Kinship and demographic behavior in the past. International studies in population*, 7 (pp. 105–134). Dordrecht: Springer. doi: [10.1007/978-1-4020-6733-4\\_5](https://doi.org/10.1007/978-1-4020-6733-4_5)
- Vulsma, R. F. (2002). *Burgerlijke stand en bevolkingsregister*. Den Haag: Centraal Bureau voor de Genealogie.
- Walhout, E. (2019). *An Infants' Graveyard? Region, religion and infant mortality in North Brabant, 1840-1940* (PhD Thesis Tilburg University). Retrieved from [https://pure.uvt.nl/ws/portalfiles/portal/29027270/Walhout\\_An\\_Infants\\_25\\_01\\_2019.pdf](https://pure.uvt.nl/ws/portalfiles/portal/29027270/Walhout_An_Infants_25_01_2019.pdf)
- Walhout, E., & van Poppel, F. (2003). “De vermelding des beroeps: eene ijdele formaliteit?": Twee eeuwen vrouwelijke beroepsarbeid in Nederlandse huwelijksakten. *Tijdschrift voor Sociale Geschiedenis*, 29(3), 301–332.
- Woods, R. I., Løkke, A., & van Poppel, F. (2006). Two hundred years of evidence-based perinatal care: Late-fetal mortality in the past. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 91(6), 45–447. doi: [10.1136/adc.2006.098178](https://doi.org/10.1136/adc.2006.098178)
- Yeung, G. Y. C., van den Berg, G. J., Lindeboom, M., & Portrait, F. R. M. (2014). The impact of early-life economic conditions on cause-specific mortality during adulthood. *Journal of Population Economics*, 27(3), 895–919. doi: [10.1007/s00148-013-0497-1](https://doi.org/10.1007/s00148-013-0497-1)

- Zijdeman, R. L. (2009). Like my father before me: Intergenerational occupational status transfer during industrialization (Zeeland, 1811–1915). *Continuity and Change*, 24(3), 455–486. doi: [10.1017/S0268416009990166](https://doi.org/10.1017/S0268416009990166)
- Zijdeman, R. L. (2010). *Status attainment in the Netherlands, 1811–1941. Spatial and temporal variation before and during industrialization* (PhD thesis Universiteit Utrecht).
- Zijdeman, R. L., & Corten, R. (2012). De keuze van de huwelijksgetuige: Sterke en zwakke bindingen in gemeentelijke context (Nederland, 1851–1935). In B. Völker (Ed.), *Over gaten, bruggen en witte paters. Sociaal kapitaal in sociologisch onderzoek* (pp. 171–183). Amsterdam: Rozenberg Publishers.
- Zijdeman, R. L., & Lambert, P. S. (2010). Measuring social structure in the past: A comparison of historical class schemes and occupational stratification scales on Dutch 19th and early 20th century data. *Journal of Belgian History/Belgisch Tijdschrift voor Nieuwste Geschiedenis/Revue Belge de Histoire Contemporaine*, 40(1–2), 111–141. Retrieved from <https://www.journalbelgianhistory.be/nl/journal/belgisch-tijdschrift-voor-nieuwste-geschiedenis-2010-1-2/measuring-social-structure-past>
- Zijdeman, R. L., & Maas, I. (2010). Assortative mating by occupational status during early industrialization. *Research in Social Stratification and Mobility*, 28(4), 395–415. doi: [10.1016/j.rssm.2010.06.004](https://doi.org/10.1016/j.rssm.2010.06.004)
- Zijdeman, R. L., & Mandemakers, K. (2008). De rol van het gymnasium en middelbaar onderwijs bij de intergenerationele overdracht van status, Nederland 1865–1940. In I. Maas, M. H. D. van Leeuwen & K. Mandemakers (Eds.), *Honderdvijftig jaar levenslopen. De Historische Steekproef Nederlandse bevolking* (pp. 149–172). Amsterdam: Amsterdam University Press.



HISTORICAL LIFE COURSE STUDIES  
VOLUME 9 (2020), published 10-09-2020

# Historical Chinese Microdata

## 40 Years of Dataset Construction by the Lee-Campbell Research Group

Cameron D. Campbell

Hong Kong University of Science and Technology & Central China Normal University

James Z. Lee

Hong Kong University of Science and Technology

### ABSTRACT

The Lee-Campbell Group has spent forty years constructing and analysing individual-level datasets based largely on Chinese archival materials to produce a scholarship of discovery. Initially, we constructed datasets for the study of Chinese demographic behaviour, households, kin networks, and socioeconomic attainment. More recently, we have turned to the construction and analysis of datasets on civil and military officials and other educational and professional elites, especially their social origins and their careers. As of July 2020, the datasets include nominative information on the behaviour and life outcomes of approximately two million individuals. This article is a retrospective on the construction of these datasets and a summary of their findings. This is the first time we have presented all our projects together and discussed them and the results of our analysis as a single integrated whole. We begin by summarizing the contents, organization, and notable features of each dataset and provide an integrated history of our data construction, starting in 1979 up to the present. We then summarize the most important results from our research on demographic behaviour, family, and household organization, and more recently inequality and stratification. We conclude with a reflection on the importance of data discovery, flexibility, interaction and collaboration to the success of our efforts.

**Keywords:** China, Historical demography, Historical big data, Quantitative history, Inequality

DOI article: <https://doi.org/10.51964/hlcs9303>

© 2020, Campbell, Lee

This open-access work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/), which permits use, reproduction & distribution in any medium for non-commercial purposes, provided the original author(s) and source are given credit. See <http://creativecommons.org/licenses/>.

## 1 INTRODUCTION

For over forty years, beginning in 1979, the Lee-Campbell Group has devoted considerable effort to locate, construct, and analyse individual-level datasets based largely on Chinese archival materials to produce a scholarship of discovery. Initially, we studied Chinese demographic behaviour, households, kin networks, and socioeconomic attainment. We constructed datasets that followed individuals from birth to death and families and households across multiple generations. More recently, we have turned to the study of the social origins and careers of civil and military officials and other educational and professional elites. We are now constructing datasets that describe the civil service careers of Qing dynasty officials from the 18th century to the beginning of the 20th century, the social origins and educational trajectories of university students during the 20th century, the qualifications and careers of government officials and educated professionals largely during the Republican era (1911–1949), and the experience of hundreds of thousands of Chinese peasants and their families during the process of rural reconstruction from land reform in the mid-1940s to Peoples Communes in the mid-1960s. Based on these data, we have published seven academic books and some 70 academic articles, mostly in English, eleven of which have won thirteen best academic prizes or equivalent recognition, including four books and two articles we wrote in English and two books and three articles we wrote in Chinese.<sup>1</sup>

This article is a retrospective on these projects and a summary of their findings. In part one, we overview the datasets themselves, summarizing their contents, organization, and notable features. In part two, we provide an integrated history, starting in 1979 with James Lee's effort to locate systematic historical demographic microdata in China, and continuing up to the present. In part three, we summarize the contributions from the analysis of these datasets beginning with the key demographic outcomes that were the focus of our early work, and then move to inequality and stratification. We conclude with reflections based on our experience.

This is the first time we have presented all our projects together and discussed them and the results of our analysis as a single integrated whole. By doing so, we clarify the full scale and range of our efforts for readers who may only be familiar with some of the projects. The projects share an emphasis on the discovery of social phenomena by an inductive approach that prioritizes careful description, sensitivity to the institutions that produced the sources, and awareness of the historical and social context in which the individuals and families we study are embedded. Up-to-date information on each project is available at our group website.<sup>2</sup>

## 2 DATASETS

In this section we present the basic features and current status of each dataset. For this introduction, we organize our datasets into four categories: 1) family, kinship and demographic behaviour, 2) education, 3) employment, and 4) rural reconstruction. Each category includes a variety of datasets, some largely complete, and some still in progress. The populations are all Chinese, largely from the 18th, 19th, and 20th centuries. As of July 2020, the datasets include 8,167,457 records with nominative information on the behaviour and life outcomes of 1,753,700 individuals who were the focus of those records

---

1 See our group web page at <https://www.shss.ust.hk/lee-campbell-group/publications/> for a list of these award-winning publications and the datasets with which they are associated. Twelve of the authors are active or former group members: Cameron Campbell, Bijia Chen, Shuang Chen, Hao Dong, James Z. Lee, Lan Li, Chen Liang, Bamboo Y. Ren, Danching Ruan, Feng Wang, Emma Zang, and Hao Zhang. Other co-authors are largely from the Eurasia Project in Population and Family History notably Tommy Bengtsson, Marco Breschi, Satomi Kurosu, Matteo Manfredini, Michel Oris, Noriko Tsuya, and George Alter.

2 The Lee-Campbell Group website describes ongoing projects, affiliated faculty, students, and coders, and provides links to publicly released data and documentation: <https://www.shss.ust.hk/lee-campbell-group/>.

as well as several hundred thousand related individuals.<sup>3</sup> Table 1 summarizes their contents. Three datasets and accompanying documentation are already available for download which we introduce in detail below with links and other information.

Table 1 *Lee-Campbell Group Datasets as of June 2020*

Name	Records	Persons
Family, Kinship and Population	2,975,211	488,675
Education	414,021	339,777
Employment	4,290,561	437,584
Rural Reconstruction	487,664	487,664
Total	8,167,457	1,753,700

What these diverse datasets have in common is that each seeks to record a substantively interesting well-defined target population in its entirety. This approach is different from constructing a statistically representative sample from a large target population in order to make inferences about it. By including entire target populations, we can describe broader social, political, and economic processes involving important or influential population subgroups in more detail than is possible with representative samples drawn from the general population. This is particularly important for upper-class populations such as nobles, civil servants, university faculty, and elite university graduates who may be important agents and/or harbingers of change, but account for only a small fraction of the general population and in a representative sample would include too few cases to allow much analysis of their composition, function, and change over time.

We introduce the datasets in detail below. The family, kinship and population datasets are defined by geographic area and hereditary status. Two are largely rural populations in northeast China, while the third consists of members of the Qing Imperial Lineage, almost all of whom lived in Beijing and what is now Shenyang. The education datasets include students from almost all the major Republican-era Chinese universities and from two major universities after 1949, as well as the vast majority of Chinese who graduated from foreign universities before the early 1950s. The employment datasets include nearly all civil and many military Chinese officials employed between 1760 and 1912, as well as separate nominative information for Chinese professionals from the fall of the Qing in 1912 to the early 1950s including almost all certified accountants, health professionals, engineers, and university faculty nationally as well as legal professionals in Shanghai and Beijing. Finally, the rural reconstruction datasets are drawn from nominative lists of individuals organized often by household for entire villages, rural brigades and communes, and even entire counties undergoing rural reconstruction either during Land Reform in the late 1940s or subsequent reorganization in the early and mid-1960s.

## 2.1 FAMILY, KINSHIP AND DEMOGRAPHIC BEHAVIOUR

Inspired by earlier studies on Chinese population history by Ping-ti Ho (1959) as well as by the contributions to English and French social economic history by Louis Henry (Gautier & Henry, 1958), Peter Laslett (Laslett & Wall, 1972) and their colleagues and students, our initial efforts at collecting individual level microdata for historical China focused on family, kinship and demographic behaviour. The datasets we eventually produced, collectively referred to as the China Multi-Generational Panel Datasets (hereafter CMGPD), are amenable to event-history analysis to examine community and family contextual influences on demographic behaviour and socioeconomic outcomes (Dong, Campbell, Kurosu, Yang, & Lee, 2015b). Table 2 summarizes the CMGPD datasets.

<sup>3</sup> For the family, kinship and population datasets, each record only provided information about the person who was the focus of the record, akin to an entry in a census form. For many of the remaining datasets, in addition to details about the individual who was the focus, records sometimes also listed the names and sometimes occupation or other information of related individuals, typically their spouses, parents, or other relatives.

Table 2 *Family, Kinship and Demographic Behaviour Datasets*

Name	Acronym	Time Span	Households	Records	Persons	Started
China Multi-generational Panel Datasets	CMGPD					
Liaoning	CMGPD-LN	1774–1909	209,880 <sup>a</sup>	1,513,352	266,091	1982
Shuangcheng	CMGPD-SC	1866–1913	156,711	1,346,826	107,551	2004
Imperial Lineage	CMGPD-IL	1633–1933		115,033	115,033	1990
Total				2,975,211	488,675	

<sup>a</sup> Refers to the period 1789–1913. The CMGPD-LN does not distinguish individuals by household before 1789, only by an administrative unit we refer to as household group.

The China Multigenerational Panel Database-Liaoning (hereafter CMGPD-LN) and China Multigenerational Panel Database-Shuangcheng (hereafter CMGPD-SC) are based on sets of population registers covering different administrative populations within Liaoning and Shuangcheng that in terms of format and organization resemble census-like listings of households and their members compiled at frequent intervals. Households and individuals are listed in roughly the same order in every edition, making the manual linkage of their records across time straightforward. Using the longitudinal data that we construct through record linkage, we can not only study the life histories of individuals, but also the histories of households and lineages. Because the content and organization of the CMGPD-SC and CMGPD-LN resemble those of population sources for other countries, they have been used in comparative studies, most notably in the Eurasia Project on Population and Family History introduced below.

The CMGPD-SC and CMGPD-LN both record household of residence, relationship to household head, demographic outcomes including birth, marriage, death, and basic measures of socioeconomic status. The populations they record are closed in the sense that entries and exits are rare, and when they occur, the timing is recorded, so at any given point in time the set of individuals at risk of experiencing an event is well-defined. Moreover, the target communities for each register series are recorded completely. Because the data are organized by residential household and are also multi-generational, and the detail on relationship to household head allows for children to be linked to their parents, grandparents, and other kin, they allow for us to embed individuals in their households and larger kin networks and examine how their life outcomes depend on the characteristics of distant kin and ancestors.

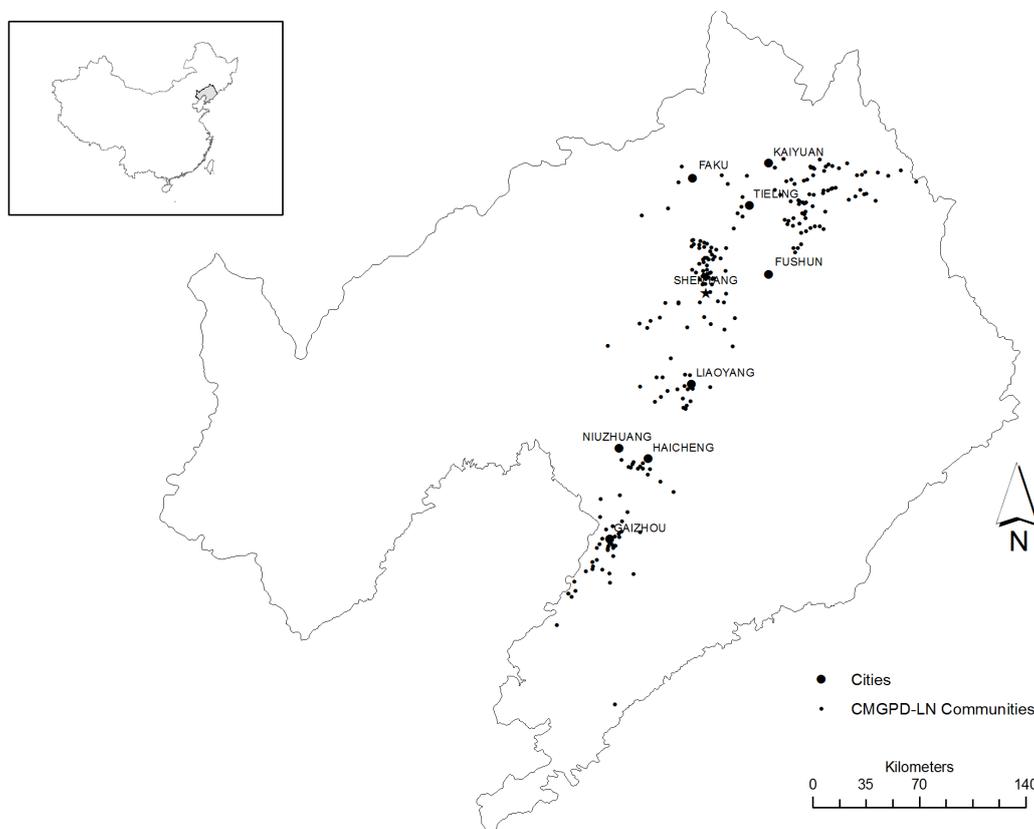
The China Multigenerational Panel Dataset-Liaoning includes 698 communities over a swath of what is now Liaoning province between 1749 and 1909 (Lee, Campbell, & Chen, 2010). The 2010 CMGPD-LN public release includes the contents of 732 population registers covering 29 administrative populations. The registers recorded each individual and household every three years. The resulting dataset includes 1.51 million records of individuals and 209,880 records of households. Through linkage of individual records, we reconstruct the histories of 266,091 individuals over as many as seven generations. We also link households over time. The data and accompanying documentation are available for download at the Inter-university Consortium for Political and Social Research (hereafter ICPSR).<sup>4</sup> Each of these included a specific administrative population living in different geographic areas.

CMGPD-LN communities are scattered over a large area roughly equivalent in size to the Netherlands (see map 1), and were economically, ecologically, and geographically diverse, including coastal communities who relied on fishing as well as farming, inland communities who cultivated fruit orchards as well as dry field agriculture, and mountain communities which supplemented such activities with hunting and gathering (Ding, Guo, Lee, & Campbell, 2004). The series include populations of regular

4 For CMGPD-LN data and documentation, please visit <https://doi.org/10.3886/ICPSR27063.v10>. In the last three years, the documentation has been downloaded 8,377 times and data has been downloaded 2,821 times (According to <https://pcms.icpsr.umich.edu/pcms/reports/studies/27063/utilization> accessed on July 14, 2020).

farmers who were tenants on state land as well as specialized populations who supplied the state in-kind levies of fish, honey, mink pelts, and other goods. Unusually for an historical Chinese source, the registers record married and widowed women completely and in detail. Like many historical Chinese sources, however, they tend to omit children who died young, especially if they were female.

Map 1 *China Multigenerational Panel Dataset-Liaoning Communities*

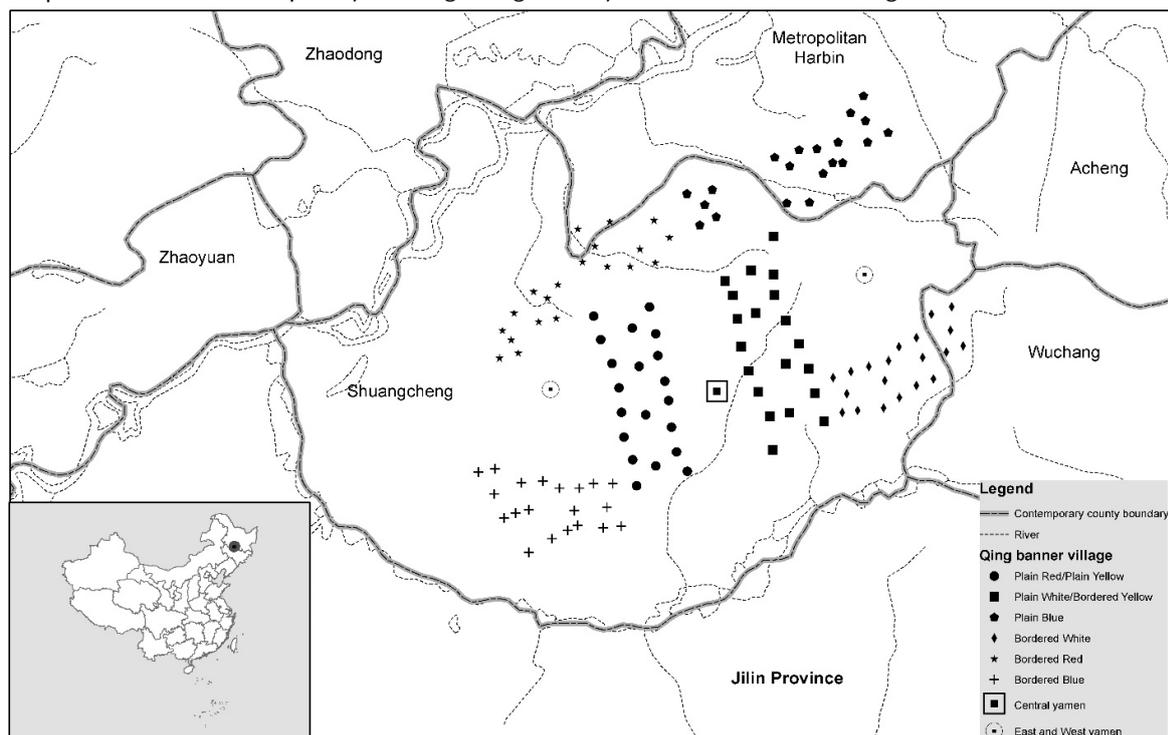


Source: Reproduced from page 4 of Lee, Campbell and Chen (2010).

The China Multigenerational Panel Dataset-Shuangcheng covers 129 communities in Shuangcheng county in Heilongjiang province from 1866 to 1913 (Chen, 2017; Wang et al., 2013). It contains 1,346,826 individual records and 156,711 household records. Through linkage, we reconstruct the histories of 107,551 individuals over as many as five generations. We also follow households over time. Map 2 presents the geographic locations of these communities and shows the location of Shuangcheng location in contemporary China. We publicly released the CMGPD-SC along with accompanying documentation at ICSPR in 2014.<sup>5</sup> The data are annual and drawn from 14 separate register series. Unlike the CMGPD-LN which covers a much larger area, CMGPD-SC communities are confined to one 3000 square kilometre county directly south of the city of Harbin. Relative to the CMGPD-LN, the data in the CMGPD-SC have more detail on their socioeconomic characteristics. Reflecting the population's diverse origins, for example, the registers record each household's official ethnicity: Manchu, Mongol, Han, Xibo and others. Moreover, by linking the CMGPD-SC household records to separate CMGPD-SC land registers, the CMGPD provides data on each household's landed wealth distinguishing between assigned and acquired land. Like the CMGPD-LN, the CMGPD-SC records widows and married women in detail. It records somewhat more children who died young and more daughters than the CMGPD-LN, but such recording is hardly complete.

5 In the last three years the documentation has been downloaded 6,240 times and data has been downloaded 2,389 times (According to <https://pcms.icpsr.umich.edu/pcms/reports/studies/35292/utilization> accessed on July 14, 2020).

Map 2 Contemporary Shuangcheng County with CMGPD-SC Villages



Source: Reproduced from Wang et al. (2013, p. 5). Original by Matthew Noellert using historical geographic coordinates provided by Yuxue Ren, Shanghai Jiaotong University Department of History, and base map data from the Harvard Yenping Institute's China Historical Geographic Information System (2007).

The China Multigenerational Panel Dataset-Imperial Lineage (CMGPD-IL) records 115,033 members of the Qing Imperial Lineage and another 135,000 or so related individuals such as spouses from before the founding of the Qing dynasty in 1644 until 1933, two decades after the fall of the Qing (Cai, Lee, Campbell, & Myers, 1994; Lee, Campbell, & Wang, 1993; Lee & Guo (Eds.), 1994). 70,000 members are sons and daughters in the main line (*zongshi*) of the lineage and the remaining 45,000 or so are sons in the collateral line (*jueluo*) (Wang, Campbell, & Lee, 2010; Wang, 2012). In contrast with most lineage genealogies in China, these records were compiled prospectively by the Imperial Lineage Office (*zongrenfu*) which throughout the Qing employed 50 to 60 officials to record Imperial Lineage members, who resided almost exclusively in Beijing and Shenyang, and administer their affairs from cradle to grave. The 28 editions of the Jade Records (*yudie*) produced by the Office between 1660 and 1921 are among the most detailed and complete records of fertility and infant and child mortality for a large Chinese population before the mid-20th century. Unlike the CMGPD-LN or CMGPD-SC they do not record residential household composition. However, they do include almost all births, including daughters, official titles and employment, notable events, and the timing of exits via death and (for daughters) out-marriage (Ju, 1994). By contrast, the privately compiled lineage genealogies that were used in many studies of Chinese historical demography rarely record daughters or wives, and tend to omit sons who died in infancy, childhood, and even early adolescence, as well as adult males who never married or married but had not surviving sons (Campbell & Lee, 2002a). By the 19th century, the Imperial Lineage was internally diverse in terms of the social and economic status, including close relatives of the emperors who had a variety of privileges, and very distant relatives of the emperors whose status was mundane.

## 2.2 EDUCATION

As our research interests moved from population and family history to long-term trends in social mobility and social stratification, we expanded our collection to include individual student records in university archives.

We collectively refer to these datasets as the China University Student Datasets (hereafter CUSD). We distinguish between overseas students (OS) who graduated from foreign universities and domestic

university students during the Republic of China (ROC) and Peoples Republic of China (PRC). Based mostly on records of matriculating university students, these datasets typically include their name, major, place of origin, current address, previous education, and the names and occupations of their parents, and sometimes guarantors. The data accordingly not only provide information on students' family origins, but also for some students allow linkage of their records to those of their parents or siblings in the CUSD as well as to their and their family members' records in other datasets.<sup>6</sup> Table 3 summarizes the CUSD.

Table 3 *Education Datasets*

Name	Acronym	Time Span	Records	Persons	Started
China University Student Datasets	CUSD				
Republic of China	CUSD-ROC	1912–1949	165,981	136,220	2010
People's Republic of China	CUSD-PRC	1949–2003	150,893	150,893	1998
Overseas Students	CUSD-OS	1847–1948			2015
In Japan			64,164	32,543	
In the United States			12,457	11,289	
In Europe			7,402	7,356	
In Soviet Union			766	758	
Other			87	81	
Not Specified			12,271	637	
Subtotal			97,147	52,664	
Total			414,021	339,777	

The China University Student Dataset-Republic of China (CUSD-ROC) covers university students in the Republican (1912–1949) era (Liang, Dong, Ren, & Lee, 2017; Ren, Liang, & Lee, 2020). It includes all or partial student registration records for 34 Republican Chinese universities. While these 34 universities represent only one-third of the universities during the Republic of China, they account for 90% of the surviving student registration records we have located in Chinese university and administrative archives. They include most of the major public, private, and missionary universities. As of January 2020, we have entered 165,981 records of 136,220 students from 34 universities. Almost all these records include student's major, age, gender, and place of origin. Most student records also include the names, occupations, and addresses of at least one parent, and in some cases of grandparents and guarantors as well. Entry of such information on parents, grandparents, and guarantors is still in progress, and we hope to add data from additional universities whose data we have already located.

The China University Student Dataset-People's Republic of China (CUSD-PRC) includes information for 64,500 undergraduate students who matriculated at Peking University between 1952 and 1999 and 86,393 undergraduate students who matriculated at Suzhou University between 1933 and 2003. The datasets are important both for their focus on elite university students in the People's Republic of China, and for their coverage of the last half of the 20th century. Peking University is one of the top national universities in China, and Suzhou is one of the best ranked regional universities. While censuses and retrospective surveys carried out since the 1980s identify college graduates, only a small number of recent surveys specify the university they attended, thus with only a few exceptions it was not previously possible to study the social or geographic origins of students at elite institutions, let alone from the 1950s to the present.<sup>7</sup>

<sup>6</sup> Entry is ongoing and we expect the number of students linked to parents or other kin to rise substantially in the coming years, thus we do not present numbers here.

<sup>7</sup> See pages 24–37 of Liang et al. (2013) for details on the construction of the CUSD-PRC, pages 37–46 for a discussion of the methods used in the analysis, and 46–57 for the contents. To protect privacy of the students in the records, entry was carried out on-site at both universities by university personnel and data remained there. Analysis of identifying data was also carried out onsite by university personnel. Analysis offsite relied on non-identifying tabulations, transformation or other calculations that had been conducted onsite.

The newest university student dataset, the CUSD-Overseas (OS), includes 52,664 Chinese students who pursued education overseas from the late 19th to the mid-20th century, accounting for 75 to 80% of the estimated 65,000–70,000 Chinese students who graduated from foreign universities during this period. As of June 2020, the dataset includes 64,164 records for 32,543 Chinese students in Japan, 12,457 records for 11,289 Chinese students in the USA, 7,402 records for 7,356 students in Europe, and over a thousand students who studied in the Soviet Union and elsewhere or for whom country of study is not available. While the CUSD-OS are based on Chinese and foreign government records of overseas Chinese students and graduates and rarely include information on students' family members, such information will be available via linkage for those students whose undergraduate records are in the CUSD-ROC.

## 2.3 EMPLOYMENT

We recently have turned to constructing a variety of large datasets on individual employment in the professions and in civil and military service in late imperial, Republican, and contemporary China. Table 4 lists our employment datasets.

Table 4 *Employment Datasets*

	Acronym	Time Span	Records	Persons	Started
China Government Employee Datasets	CGED				
Qing	CGED-Q	1760–1912	4,178,078	346,541	2013
Republic of China	CGED-ROC	1911–1949	51,806	30,365 <sup>a</sup>	2018
China Professional Occupation Datasets	CPOD				
Republic of China	CPOD-ROC	1911–1949	55,178	55,178	2016
People's Republic of China	CPOD-PRC	1949–1952	5,500	5,500 <sup>a</sup>	2015
Total			4,290,562	437,584	

<sup>a</sup> Approximate because entry and cleaning are in an early stage and still ongoing.

The largest and most developed of these datasets is the China Government Employee Dataset-Qing (CGED-Q) (Chen, Campbell, Ren, & Lee, 2020; Ren, Chen, Hao, Campbell, & Lee, 2016). The core information for the CGED-Q comes from the *jinshenlu*, a roster of civil offices that was compiled every three months during the Qing and listed almost every regular, salaried civil office and included the holder's name, place of origin, banner affiliation (if any), location of post, job title, and other details. Positions ranged from high offices in the Six Ministries and other central government units down to low-level offices in county administrations. Each edition lists 13,000–15,000 employees.<sup>8</sup> As of July 2020, we have entered 4,178,078 records of 346,541 officials for the period between 1760 and 1912.<sup>9</sup> Most of these are from the period between 1830 and 1912, during which coverage of surviving editions is nearly complete. We are releasing the data in stages. Microdata for the period 1900–1912 are already available for download at the HKUST DataSpace and at a mirror site maintained by Renmin University Institute of Qing History.<sup>10</sup>

Nominative linkage of the records of the same official in successive editions in the CGED-Q allows us to construct and study their career histories. Linkage procedures depended on whether officials

8 The CGED-Q also includes some lists of military officials from a roster *zhongshubeilan* that originally was also compiled every three months. Each edition recorded 7000–8000 military officials.

9 Of these, 3,606,301 are records of civil offices and 518,596 are records of military offices. 284,264 officials started their careers as civil officials and 60,807 officials started their careers as military officials.

10 Data and documentation were first made available at the HKUST Dataspace in May 2019: <https://doi.org/10.14711/dataset/E9GKRS>. As of July 14, 2020, data has been downloaded 741 times at the HKUST Dataspace and the CGED-Q project page at the Lee-Campbell Group website has had 14099 visits. We are collaborating with the Renmin University Institute of Qing History for the release. They maintain a mirror site to facilitate access to the data and documentation for users in mainland China. As of May 7, 2020, the Renmin University mirror site has had 3471 unique visitors.

had a hereditary affiliation with the Eight Banners by virtue of their descent from the conquest elite who established the Qing in 1644. The officials who were not affiliated with the Eight Banners were mostly Han Chinese and can be linked based on their surname, name, and province and county of origin. The combination of these four attributes was almost always unique, and the primary challenge is addressing a relatively small number of cases where records of the same individual are not linked because their name is written slightly differently in two editions, usually because one character is replaced by another that looks or sounds similar. Linkage of officials who were affiliated with the Eight Banners is more difficult because instead of a province and county of origin, their banner affiliation was recorded. Moreover, most officials who were bannermen were Manchu or Mongol and did not have surnames recorded. The primary challenge is accordingly the opposite of the one we face for non-banner officials. While approximately 86% of the combinations of name and banner affiliation are unique, for the remainder who are not unique we use additional information to prevent records of different individuals who have the same name and banner affiliation from being linked together as if they were one person. We are also entering and linking information on the family backgrounds of officials and other characteristics from records of exam degree holders and other sources.

We have also begun to extend our coverage of employment to include government officials and educated professionals in the Republican era and the early years of the People's Republic of China. The data are important for our understanding of state building and the emergence of educated professionals as a distinct social group during this period. They also allow for comparisons of officials and some professionals who served in the Qing, Republican, and early Peoples Republic of China. For the CGED-ROC (Republic of China) we have entered 31,658 records of government officials who served between 1911 and 1949. These include with some overlap 9,988 officials from the Ministry of Education, the Ministry of Defence, the Academia Sinica, and the five administrative branches called the Control, Examination, Executive, Judicial, and Legislative *Yuan*, and 21,580 officials from the Transportation and Railroad Ministries. Common variables include name, sex, age, place of origin, education credentials, current employment and employment history. Acquisition of relevant materials continues, and this dataset should expand substantially.

The China Professional Occupation Datasets-Republic of China (CPOD-ROC) date back to 2016 when Bamboo Ren located related sources from the Liaoning Provincial Archive at FamilySearch (formerly the Genealogical Society of Utah). Bamboo subsequently worked with other group members, notably Yibei Wu, in archives and libraries at Beijing, Hangzhou, Nanjing, and Shanghai, to compile five discrete datasets. Of the 55,178 currently entered, 18% are medical doctors, 36% are university faculty, and another 36% are engineers. The remaining 10% are lawyers and certified accountants. Data entry is ongoing, and we expect the number of professions and especially the number of professionals for whom we have data to increase rapidly.

## 2.4 RURAL RECONSTRUCTION

Finally, motivated largely by student interests to better understand China's rural revolution, we have collected and continue to collect nominative individual level datasets, collectively referred to as the China Rural Reconstruction Datasets (CRRD), to research China's rural reconstruction especially during the third quarter of the 20th century. Table 5 summarizes the rural reconstruction datasets.

Table 5 *Rural Reconstruction Datasets*

	Acronym	Time Span	Households	Records	Persons	Started
China Rural Reconstruction Dataset	CRRD					
Siqing	CRRD-SQ	1946–1966	25,050	63,905 <sup>a</sup>	63,905	2013
Land Reform	CRRD-LR	1946–1948	82,514	423,759	423,759	2011
Total			107,564	487,664	487,664	

<sup>a</sup> The household forms only record names and details of individuals who are over the age of 15 *sui*, that is 13.5 years Western years or older. Counts of the numbers of individuals below that age are provided on the original household form and were entered but no other details about them are available.

We do so because one of the defining features of 20th century China was the transformation of the world's largest agrarian society during this period. We are constructing two datasets that record information about individual and household experiences during the most dramatic stages of this process between 1946 and 1966, when the Chinese Communist Party carried out a nationwide redistribution of land and then gradually organized rural communities into agricultural cooperatives and ultimately People's Communes. The China Rural Reconstruction Dataset-Land Reform (CRRD-LR) was created to study the nationwide Land Reform Movement from 1946 to 1953. During this movement, local governments in many parts of rural China kept systematic records of land reform events and activities. These records include detailed individual- and household-level registers of property expropriation and reallocation and the political struggles that accompanied this redistribution of wealth. Currently the CRRD-LR contains county-wide data on the land reform experiences of over 80,000 households with approximately 400,000 individuals in Shuangcheng, Heilongjiang between 1946 and 1948.

The China Rural Reconstruction Dataset-Siqing (CRRD-SQ) is one of the most systematic and detailed sources available on social and economic change in rural China from before land reform in the 1940s up to the eve of the Cultural Revolution in 1966 (Xing, Campbell, Li, Noellert, & Lee, 2020). It is based on household social class registration forms compiled in rural areas around 1966 as part of the Socialist Education Campaign also known as the Four Clean-up (*Siqing*) Movement. The CRRD-SQ currently contains data from over 25,000 of these household forms, one quarter in collaboration with the Shanxi University Research Centre for Chinese Social History, from four provinces: Shanxi, Hebei, Inner Mongolia, and Guangdong. Each form records two to three pages of information per household, including their property holdings and occupations before and after land reform in the late 1940s, at the time when cooperatives were formed in the mid-1950s, and at the time of compilation in 1965 and 1966; the household head's social relations, a three-generation family history, and social, demographic, and political details on every household member over 15 *sui*, that is approximately 13.5 years or older according to Western ages.

## 3 HISTORY

Looking back over the last forty years, we can distinguish three distinct phases in terms of dataset construction and to some extent research. In the first phase, from 1979 to 1989, transcription and analysis were slow because of limitations in funding, technology, and support personnel. Work focused on demographic analysis of early iterations of the CMGPD-LN. In the second phase, from 1990 to 2010, data entry accelerated as stable funding became available to support a core team of full-time data coders, first in the USA and then in the People's Republic of China. Data coverage expanded to include the entire current CMGPD as well as the CUSD-PRC. In the third phase, from 2010 to the present, the range of population categories broadened to include government officials, professionals and other educated elites with the initiation of the CGED, the CPOD, and the CUSD-ROC and CUSD-OS, as well as such topics as rural reconstruction with the creation of the CRRD-LR and CRRD-SQ. Here we narrate progress across these three phases. The emphasis is on when and how each project were initiated, the participants and their contributions, and key transitions in terms of approach and scale.

### 3.1 PHASE 1 — THE BEGINNING, 1979–1989

We began more than forty years ago when James Lee began to look for quantizable nominative individual level microdata in historical archives in mainland China beginning with a winter-long visit to the First Historical Archives in Beijing in 1979. Inspired by the quantitative historians and social scientists who in the 1960s and 1970s transformed our understanding of family and population in past times in Europe and North America by the construction and analysis of datasets from archival sources, he hoped to do the same for historical China.<sup>11</sup>

11 Funding during this decade came from a combination of internal Caltech resources as well as from external support from the National Endowment for the Humanities and from a Wang Fellowship in Chinese Studies. In addition, support from the Academia Sinica and the National Program for Advanced Study and Research in China funded nearly all our travel as well as some of our research expenses in Beijing, Shenyang and Taipei.

In 1982 at the advice of Deyuan Ju, Lee visited the Liaoning Provincial Archives and obtained microfilms of five household registers from Daoyi covering the period between 1774 and 1798.<sup>12</sup> Together with Robert Eng, an economic historian with prior experience with Japanese historical population registers, Lee developed a coding scheme and personally transcribed the contents of the 1774, 1780, 1786, and 1792 registers into a fixed-column format, first on paper forms and then into digital files.<sup>13</sup> Lee also took a course in demographic methods in 1984 offered by the Graduate Group in Demography at the University of California, Berkeley. Working primarily with various California Institute of Technology (Caltech) undergraduates, he published the first analyses of Chinese mortality, fertility, and household structure for specific historical populations on the Chinese mainland before the 20th century using household register microdata (Lee, Anthony, & Suen, 1988; Lee & Eng, 1984; Lee & Gjerde, 1986).<sup>14</sup> He also published with William Lavelly and Feng Wang an influential article on how new historical and contemporary microdata were reshaping the understanding of Chinese demographic behaviour (Lavelly, Lee, & Wang, 1990).

While Lee acquired additional 19th century household and population registers for Daoyi beginning in 1985, research using such microdata did not advance significantly until the arrival in the summer of 1987 of Cameron Campbell, a Caltech sophomore (second year student) majoring in Electrical Engineering with a side interest in Chinese history he had developed in high school.<sup>15</sup> Campbell had prior training and experience with database programming. After going over the various C programs that had been written for specific data transformations and calculations for the Daoyi registers, he proposed to Lee a new workflow where the data would be managed in dBase III+ (later dBase IV) and then exported to SPSS for analysis. Campbell began to develop the new code in summer 1987 when he and Lee re-visited the First Historical Archives in Beijing and the Liaoning Provincial Archives in Shenyang.<sup>16</sup> Processing included construction of flag variables for the occurrence of demographic events, identifiers for records of the same individual in different registers and links between kin, measures of household structure and composition, and measures of context at the individual level including the presence and absence of specific kin. This simplified data entry and created new possibilities for analysis to move beyond the calculation of rates and proportions.<sup>17</sup>

### 3.2 PHASE 2 — ACCELERATION, 1990–2009

Entry of the CMGPD-LN subsequently accelerated, first in 1990 thanks to support from the Academia Sinica and the National Science Council in Taiwan, and again in 1999 thanks to a large private gift to James Lee. Before 1990, only one of the 29 administrative populations that eventually made up the CMGPD-LN, Daoyi, had been entered, along with a few registers from another administrative population, Gaizhou, yielding about 100,000 observations. In contrast, between 1990 to 1999, we

12 Additional registers were filmed at the Liaoning Provincial Archives in 1985 and 1987 that extended the series to 1873.

13 See Lee and Campbell (1997, pp. xix–xxi) for complete lists of the people who helped with the entry and analysis of the Daoyi registers, the coders who entered the data, and the funding sources.

14 By this time, Arthur Wolf had already done related work for Taiwan using Japanese colonial records from the first half of the 20th century (Wolf & Huang, 1980).

15 Campbell's participation in summer 1987 and 1988 was supported by Caltech Summer Undergraduate Research Fellowships.

16 To carry out programming and analysis while in Beijing and Shenyang in summer 1987, Lee and Campbell brought a large, very heavy portable Compaq microcomputer with them from the United States. We however forgot that voltage transformers were not widely available in China and therefore in Shenyang had to rely on one that had been hand-built and which was turned on and off with a knife-switch.

17 By the time Campbell graduated in June 1989, he had switched from Electrical Engineering to a double major in History and Engineering and Applied Science. After graduation, he went to Taipei and Beijing with funding from the Watson and Durfee Foundations to learn Chinese and then spend time in Daoyi. In Taipei he spent six months studying Chinese intensively at the Mandarin Training Centre at National Taiwan Normal University. He spent the summer of 1990 working with James Lee in Taipei. In fall 1990, he went to the University of Pennsylvania to study for a PhD in demography. He spent another year in Taipei for further intensive training in Chinese in 1992–1993 at the Inter-University Program in Chinese Language Studies at National Taiwan University and then spent summer 1993 conducting dissertation research in Beijing. He was supported by a Social Science Research Council International Pre-Dissertation Research Fellowship.

entered an additional 400,000 records from 8 more administrative populations. Our increased speed was partly due to the acquisition of most of the historical household registers and related materials from the Liaoning Provincial Archives by the Genealogical Society of Utah who made these sources available to us, and partly because the increase in funding enabled us to support a larger team of coders to transcribe these data into digital datasets.<sup>18</sup> Almost all the data entry was done by coders in the United States, though two CMGPD-LN series were completed in Taiwan. In 1999, we moved data entry to mainland China where we were fortunate to have the support of one and soon afterwards three reliable and enthusiastic full-time coders: Xing Xiao, Huicheng Sun, and Jiyang.<sup>19</sup> Over the next four years they entered the remaining 19 CMGPD-LN datasets, yielding an additional 1 million observations, and subsequently spent a full person-year largely in 2010 to clean the entire 1.5 million observation CMGPD-LN in preparation for public release of these data.<sup>20</sup>

To conduct more detailed analyses of fertility and infant and child mortality than was possible with the CMGPD-LN, Lee collaborated with Huimin Lai and Sufen Liu at the Academia Sinica in Taiwan to construct the CMGPD-IL beginning in 1990. Deyuan Ju at the First Historical Archives had earlier introduced Lee to the collections of nominative historical demographic microdata from the Office of the Imperial Lineage including the Jade Registers (*yudie*) or Imperial Lineage Genealogy (Ju, 1994). Recognizing that the nearly complete recording of male and female births and infant and child deaths made the *yudie* an invaluable complement to the CMGPD-LN, which recorded few daughters, and omitted some sons who died in infancy or childhood, Lee obtained copies of these data in 1985 when the First Historical Archives in Beijing microfilmed the *yudie* for the Genealogical Society of Utah. Lee oversaw the coding of the members of the Main Line (*zongshi*) in collaboration with Huimin Lai and Sufen Liu between 1990 and 1992. Lee also recruited Feng Wang in 1989 to participate in the analysis of these data and published with Campbell an introduction to the CMGPD-IL dataset (Lee, Campbell, & Wang, 1993).<sup>21</sup> Campbell used the CMGPD-IL in his PhD dissertation which studied long-term trends in mortality in Beijing by comparing mortality patterns in the CMGPD-IL in the 18th and 19th century with mortality patterns in Beijing in the 1920s and 1930s and after 1949.<sup>22</sup> Later, working with Linlan Wang, a sociology PhD student of Lee's at Peking University, Lee added the sons from the collateral line (*jueluo*) recorded in the 1933 *Aixin Jueluo Genealogy* for her 2012 PhD thesis (Wang, 2012; Wang, Lee, & Campbell, 2010).<sup>23</sup>

In 2003, Lee who was now at the University of Michigan, began to work with Shuang Chen, a PhD student in history, to construct the CMGPD-SC to examine the relationship between landholding and demographic behaviour in Shuangcheng, Heilongjiang (Chen, 2009; Wang et al., 2013).<sup>24</sup> In addition to landholding, the CMGPD-SC allows for other analyses not possible with the CMGPD-LN, for example comparisons by registered household ethnicity. The Genealogical Society of Utah had acquired 338 population registers and 23 land registers for Shuangcheng and in fall 2003 made them available to us. Xing Xiao, Huicheng Sun, and Jiyang entered the 1.3 million records in the CMGPD-

18 Upon completion of the CMGPD-LN, we provided a name list linking records to their pages in the original sources to the Genealogical Society of Utah (now [FamilySearch](#)) for inclusion in their online search facility.

19 Another logistical challenge after moving data entry to China was processing payments to the coders. We are grateful to staff at University of Michigan ICPSR, UCLA CCPR and HKUST for their oversight and management of the grants, especially Ruth Danner at ICPSR, Lucy Shao at UCLA CCPR, and Freda Ching at HKUST.

20 Between 1999 and 2006, we also visited some 57 Liaoning villages during which we acquired almost 250 related data sets such as family genealogies.

21 Lee, Campbell and Wang (1993, p. 361) lists the scholars at the Academia Sinica who facilitated this work, the coders who entered the data, the Caltech undergraduates who helped with programming, and the funding sources. Lee and Guo (Eds.) (1994) provides a detailed history in Chinese of the project, including the data entry process and the transformation of the data to prepare it for analysis.

22 Campbell collected the mortality data for 20th century Beijing while doing archival and library research there in summer 1993. While conducting this research he was assisted by Jennifer Huang Bouey.

23 Unlike the Imperial Lineage Genealogy but like almost all other historical Chinese genealogies, the Aixin Jueluo Genealogy did not report daughters. Data entry was done principally by Jiyang with the assistance of Xing Xiao from April 2008 to July 2011.

24 Wang et al. (2013, p. viii) provides a complete list of the individuals who contributed in various ways to the construction of the CMGPD-SC.

SC between 2004 and 2007.<sup>25</sup> By contrast, entry of the 1.5 million records in the CMGPD-LN had taken approximately twenty years. Shuang Chen, who is now an associate professor of history at the University of Iowa, oversaw the data coding and led the analysis of the CMGPD-SC for her dissertation (Chen, 2009) and book (Chen, 2017) as well as for our contributions to the Eurasia Project fertility volume (Chen, Campbell, & Lee, 2014).

Our capacity to manage and analyse steadily larger datasets improved as a result of additional changes in the workflow. The dBase programs that read in the files provided by the coders to produce the files to be analysed in SPSS and later STATA were maintained by Chris Myers in the early 1990s and then again by Cameron Campbell. However, the programs were slow. In the late 1980s, when there were fewer than 70,000 records to deal with, the dBase programs that transformed the raw data entered by coders into files to be used in analysis could run for more than a day and were prone to crashing.<sup>26</sup> Subsequent improvements in processing and disk speed were offset by increases in the number of records to be handled. Finally, in the mid-1990s, we froze development of the dBase programs. They continued to be used to process incoming files and prepare a file for analysis in STATA, but they were not further developed. Creation of new variables was done in STATA. Eventually, Campbell retired the dBase programs completely and wrote STATA code to handle the entire process of importing files, organizing them, creating variables for analysis, and carrying out analysis. This reduced the time required to go from the raw files provided by coders to the work files used in analysis to just a few hours.

To better understand the context of the communities that were recorded in the CMGPD-LN and learn more about the histories of the families it recorded, we conducted fieldwork in rural Liaoning. Between 1999 and 2006, we made eight field trips to Liaoning accompanied by Gao Jing and colleagues from the Liaoning Provincial Gazetteer Office and local Gazetteer Offices during which we visited 57 largely rural communities. We spent some 250 person-days visiting descendants of the CMGPD-LN populations and collecting local sources such as genealogies, tomb inscriptions, deeds, and other family documents on these populations. We also gathered oral histories and collected information on the families during the time from the end of the CMGPD-LN in 1911 up to the time of our visit. We compared these local data with the state household and population registers in Campbell and Lee (2002a) and Ding, Guo, Lee, and Campbell (2004). In each community, we shared with the families we visited genealogies of their lineages generated from the CMGPD-LN. Many families had lost their genealogies or had only rudimentary genealogies that listed only the generation and given names of male lineage members, and the materials we provided helped them reconstruct their family histories, including names and other information on ancestors who held office or had other achievements or recognition.

Taking advantage of advances in technology, we began to use event-history analysis and other regression-based approaches to study associations between individual demographic behaviour and outcomes and household and community context. The capacity of the personal computers that we were using for processing and analysis improved dramatically. In the early 1990s, computations involving the 100,000 or so records in the Daoyi series that involved anything more than tabulation or linear regression took fifteen minutes to an hour, depending on the number of observations included, the number of variables, the type of model, and the number of models. By the late 1990s, more advanced estimations run on much larger numbers of records took much less time. By the late 2000s, calculations involving the combined CMGPD-LN and CMGPD-SC files, nearly 3 million records, could be completed in minutes on a personal computer.

Our shift to event-history analysis was also spurred by Akira Hayami's 1993 invitation to James Lee to participate in the Eurasia Project in Population and Family History, an international comparative project that studied interactions between community context, household organization, and demographic

25 During this time, our work received support from NIH NICHD 1R01HD045695-01A2 (Demographic Responses to Community and Family Context, James Lee PI).

26 While Campbell was still at Caltech, the programs were so slow and prone to crashing that when he and Lee processed the dataset to integrate newly coded registers or reprocessed the dataset to add new variables, he often slept on the floor in Lee's office, waking up every few hours to check that the programs were still running, and if they had failed, correcting the problems and restarting them.

behaviour in past times.<sup>27</sup> The project began in 1994 and yielded three volumes in a dedicated series published by MIT Press under the editorship of Lee, Bengtsson, and Alter. The first was on mortality (Bengtsson, Campbell, Lee et al., 2004), the second was on fertility (Tsuya, Wang, Alter, Lee et al., 2010), and the third was on marriage (Lundh, Kurosu et al., 2014).<sup>28</sup> Teams of researchers who had household register data from communities in Belgium, China, Italy, Japan, and Sweden collaborated to specify event-history models of mortality, fertility and marriage that could be estimated in all of the datasets which would yield results that could be compared. We discuss findings from the project below.

In the early 2000s we began to plan for the public release of the CMGPD. Lee's affiliation with the Inter-university Consortium for Political and Social Research (ICPSR) after his move to Michigan in 2003 was crucial. Conversations with Myron Guttman and others convinced us that public release of the CMGPD was not only important but, with support from ICPSR, also entirely feasible. ICPSR provided administrative support for the submission of funding applications to the National Institutes of Health and management of the resulting grants, and staff support for the deposit of CMGPD data and documentation through the ICPSR Data Sharing for Demographic Research (DSDR) program.<sup>29</sup> After James Lee moved to the Hong Kong University of Science and Technology (HKUST) in 2009, Cameron Campbell at University of California, Los Angeles (UCLA) received administrative support from the California Centre for Population Research for submission and management of a grant to support public release of the CMGPD-SC, but the release itself was again through ICPSR DSDR with support from ICPSR staff, most notably Susan Hautaniemi Leonard. From 2011 to 2014, Campbell conducted training workshops at Shanghai Jiao Tong University every summer to introduce prospective users to its contents and organization and demonstrate advanced operations to manage and analyse the data. Yuxue Ren hosted him and Dan Xu provided logistical support. Based on experiences from the workshops, Campbell, Hao Dong and James Lee produced the *CMGPD Training Guide* (Campbell, Dong, & Lee, 2013) as a companion to the *User Guides* (Lee, Campbell, & Chen, 2010; Wang et al., 2013).

In the late 2000s, we initiated a new line of work on social mobility, stratification, and inequality. We had originally considered father-son associations in social and economic outcomes in Daoyi in Lee and Campbell (1997, pp. 196–214). When Campbell first started as an assistant professor in sociology at (UCLA) in 1996, he was still focused on demography, but interactions with Donald Treiman, William Mason, Robert Mare, Judith Seltzer, Ken Sokoloff, Jean-Laurent Rosenthal and other colleagues in sociology and economics inspired a desire to take advantage of the distinctive properties of the

27 See the acknowledgments in Bengtsson, Campbell, Lee et al. (2004, pp. xi–xii), Lundh, Kurosu et al. (2014, pp. xxiii–xxv), Tsuya, Wang, Alter, Lee et al. (2010, pp. xxiii–xxv) for a history of each volume and Lundh, Kurosu et al. (2014, pp. xvii–xxi) for a recap and reflection. Campbell (2016) is a personal reminisce and reflection on his involvement in the project. See Lee and Steckel (2006) for an appreciation and appraisal of Bengtsson et al. (2004). Daniel Little has discussed the Eurasia Project in Population and Family History in a post on his blog *Understanding Society*: <https://understandingsociety.blogspot.com/2014/08/eurasia-project-on-population-and.html>.

28 We would also like to thank the many participants who discussed these books at authors-meet-critics sessions at the Social Science History Association in 2004, 2010, and 2015 and the Population Association of America meetings in 2005 and 2011, and at the summary discussion of the Eurasia Project in Population and Family History at the 2014 American Sociological Association. Among the many participants at these sessions were Douglas Anderton, Jason Beckfield, Hilde Bras, Andrew Cherlin, Jack Goldstone, David Hacker, Michael Haines, Charles Hirschman, Jan Kok, Ronald Lee, Daniel Little, Deirdre McClosky, Myron Gutmann, Karen Oppenheim Mason, Richard L. Steckel, Jan Van Bavel, and Andreas Wimmer. Before the publication of each volume, work in progress was presented at dedicated sessions at Social Science History Association, European Social Science History Conference, World Economic History Congress, and other meetings. These sessions are listed in the acknowledgments of the respective volumes.

29 NIH NICHD 1R01HD057175-01A (The Liaoning Multi-Generational Panel Dataset: Public Release and User Training, Lee and then Leonard PI) supported release of the CMGPD-LN between 2009 and 2012. NIH NICHD 1R01HD070985-01 (Multi-generational Demographic and Landholding Data: CMGPD-SC Public Release. Campbell PI) supported public release of the CMGPD-SC between 2012 and 2016.

CMGPD-LN to study social mobility, inequality, kinship and other topics.<sup>30</sup> We moved past father-son associations to study associations of socioeconomic outcomes with the characteristics of progressively wider networks of kin, starting with siblings, uncles and grandparents in the same household, then moving to kin outside the household, and then eventually to lineages. We faced a constraint, however, in that the CMGPD datasets only recorded official positions held by adult males. They provide no information about other non-agricultural occupations, and only the CMGPD-SC recorded landholding.

Inspired by this growing interest in inequality and social mobility and a desire to move beyond the study of demography and after having devoted over two decades to the collection of individual level information on socio-economic attainment and related demographic behaviour in pre-20th century China, we turned our attention to the construction of datasets for the study of inequality, social mobility, and social change in historical and contemporary China. Lee and collaborators initiated the CUSD-PRC when Lee learned of the student registration cards held in the Peking University archives. Actual entry of data at Peking University began in 2003 under the supervision of Lee's collaborators, Danching Ruan and Shanhua Yang, and Yang's PhD student Hao Zhang. Entry of data at Suzhou University began in 2006, supervised by Chen Liang, a recent Tsinghua University history PhD, who went on to a postdoc at Michigan mentored by James Lee, and is now a history professor at Nanjing University.

### 3.3 PHASE 3 — THE EXPANSION, 2009–PRESENT

The most recent phase began when James Lee moved to the Hong Kong University of Science and Technology in 2009. Campbell moved from UCLA and joined him in 2013. Proximity to the mainland allowed for frequent travel there to interact with researchers, present results, and explore new sources of systematic information on China's historical and recent past. This led to the discovery of new materials that could be transformed into databases and which allowed us to move from the study of household organization and demographic behaviour to the study of educated elites. Generous and stable intramural and extramural funding supported a further expansion of coding efforts.<sup>31</sup> It was also easier to recruit, support, and work with graduate students with relevant interests than was the case while Lee and Campbell were in the United States. Since Lee's arrival at HKUST, we have trained and collaborated with Matt Noellert, Hao Dong, and Bijia Chen who are now respectively tenured faculty in economic history at Hitotsubashi University, tenure-track faculty in sociology at Peking University, and postdoctoral fellow in history at Renmin University's Department of History. As of the summer of 2020, our current PhD students are Xiangning Li and Bamboo Y. Ren at the Hong Kong University of Science and Technology, Qin Xue at Central Normal University, and Li Yang and Yibei Wu at Shanghai Jiaotong University.

We now construct datasets for the study of inequality, social mobility, and social change in historical and contemporary China. In 2010, Chen Liang proposed a study of social origins of university students in the first half of the 20th century that would extend on the CUSD-PRC by construction and analysis of a dataset (CUSD-ROC) based on student registration cards held in historical archives throughout China (Liang et al., 2017; Ren et al., 2020). Working with others in the Lee-Campbell Group, he located the student registration cards for half of the 34 universities that currently make up the CUSD-ROC and arranged for most of their data entry. Bamboo Y. Ren, James Lee, and Mingyu Zhang located and coded the other half of these data. The process for transcription of these records differed from that of the CMGPD datasets. Rather than having a dedicated team transcribe the contents of scans of the original sources, data was entered on-site in the archives by personnel recruited locally for the task. Coding additional variables or checking the originals to resolve inconsistencies required return trips to the archives.

The next project we initiated was the CGED-Q. Cameron Campbell conceived of it as a resource for the study of Qing officialdom and the career dynamics of Qing officials in 2013 when Yuxue Ren, a postdoctoral fellow of Lee's at the University of Michigan and now an associate professor of history

30 In the late 1990s and through the 2000s, three seminar series at UCLA were especially important in inspiring Campbell to move beyond traditional historical demography: the Von Greppe Workshop in Economic History, the UCLA/RAND Joint Labour and Population Workshop, and the California Centre for Population Research Seminar Series.

31 Our work has been supported by a series of grants from the Hong Kong Research Grants Council General Research Fund: 642911 (Lee PI), 640613 (Lee PI), 16400114 (Campbell PI), 16400714 (Lee PI), 16602315 (Lee PI), 16600017 (Campbell PI), 16602117 (Lee PI), 16601718 (Campbell PI).

at Shanghai Jiao Tong University, showed him and James Lee work she was doing with records of officials in northeast China she had transcribed from quarterly *jinshenlu* in a collection of 206 editions published by the Tsinghua University Library.<sup>32</sup> Campbell, Lee, and Ren developed a plan to enter all of the 2.8 million records in this collection and an additional 1.2 million records from *jinshenlu* editions held elsewhere. This was completed in summer 2020. The coders we had relied on for the CMGPD began entering data in 2014. In 2016 we added new coders and the pace of entry doubled. Bijia Chen joined the project in 2015 while she was an MPhil student in Social Science at HKUST. She played a key role in the coordination of the data entry and then wrote her PhD dissertation on the careers of Qing officials (Chen, 2019).

While the CMGPD, CGED, and much of the CUSD were all initiated either by James Lee or Cameron Campbell, working with other senior members of our research team such as Chen Liang, our most recent data sets, on China's rural revolution (CRRD-LR and CRRD-SQ) and the rise of China's professionals (CPOD) are the initiative largely of our younger team members who found these sources and organized the data construction for their PhD theses. Matthew Noellert discovered the materials that became the basis of the CRRD-LR while conducting fieldwork in Shuangcheng in 2011 and used these data to write his 2014 PhD dissertation and 2020 book.<sup>33</sup> Moreover, while Long Xing, the Director of the Shanxi University Research Centre for Chinese Social History, initiated the China Rural Reconstruction Dataset-Siqing (CRRD-SQ) based on his collection of 7800 rural household social class registration forms, it was Noellert, working together with Yingze Hu, who oversaw the initial coding of the CRRD-SQ in 2015<sup>34</sup>, and Noellert together with HKUST PhD student, Xiangning Li, who between 2016-2019 expanded the CRRD-SQ from one to four provinces, and from eight thousand to 25 thousand households.

Similarly while the CUSD-ROC and CUSD-OS were initiated by Chen Liang in 2010 and James Lee in 2019, it was Bamboo Ren who initiated and coordinated the construction of the various CPOD datasets beginning in 2018 and Yibei Wu and Li Yang working under Bamboo's direction who coordinated the data entry of the CUSD-OS, and Yibei Wu who located the data sources and oversaw the data transcription of the CGED-ROC.

## 4 FINDINGS

We organize our review of findings by topic. We begin with studies of demographic behaviour and household organization. We follow the development of these studies from estimates of demographic rates and household structure to examinations of the implications of family hierarchy based on patterns of differentials according to household context and finally to analyses of assortative mating, relationships between household context and health and mortality in later life, and other topics. We then present work on intergenerational social mobility and inequality more generally. We start with the earliest studies of the associations between fathers' and sons' socioeconomic outcomes, move to multi-generational studies that considered the role of kin other than the father in shaping individual outcomes, and conclude with recent studies that move beyond the individual to consider kinship networks and descent groups as units of analysis, and stratification and inequality more generally. Third, we summarize recent published work on the geographic and social origins of university students in 20th century China. Finally, we summarize recent published studies of the careers of government officials during the Qing.

---

32 See Chen et al. (2020) for an English language introduction and Ren et al. (2016) for a Chinese language introduction to the project, sample of results, and complete list of the collaborators, students, and coders who contributed to the CGED-Q.

33 Huicheng Sun, Xing Xiao, and Jiyang coded the CRRD-LR largely during 2012 and linked individuals across various CLRD event registers in 2013.

34 The resulting dataset is held at Shanxi University and analysis is carried out by personnel there. As with the CUSD-PRC, any analysis we conduct is based on tabulations or other calculations produced in response to our requests, not on the original data. Since 2016, we have located and entered another 18,000 forms. Matt Noellert and Xiangning Li lead the analysis of these materials.

## 4.1 DEMOGRAPHIC BEHAVIOUR

The earliest line of work described trends and patterns in mortality, fertility, population age and sex composition, and household structure by presentation of aggregated rates and proportions. Based on five registers from Daoyi between 1774 and 1798, Lee and Eng (1984) introduced the data and presented descriptive results on birth and death rates, population age composition, and household structure. Among other findings, they showed that these sources recorded adult males and married and widowed females completely but omitted many sons who died in infancy or early childhood along with most daughters. Lee, Campbell, and Wang (1993) introduced the CMGPD-IL and presented time trends and age patterns of mortality for the members of the Imperial Lineage. Lee, Anthony, and Suen (1988) and Lee, Campbell, and Anthony (1995) showed that levels and patterns of mortality in Daoyi resembled those in other historical populations. Lee and Gjerde (1986) compared household forms in Daoyi with those in Norway and the United States to show that existing schemes for classifying household structure were inadequate and proposed a new classification scheme more amenable to comparison between Europe and other societies. Comparison to the CMGPD-LN revealed that adult Imperial Lineage males had higher death rates than adult males in Daoyi in rural Liaoning, presumably because being restricted to living in Beijing subjected them to an 'urban penalty'. Results from these early studies led to lines of research exploring relationships of fertility, mortality, and other aspects of demographic behaviour to other social and economic variables which took advantage of the individual-level detail in the data. We describe these lines of research below.

### 4.1.1 FERTILITY

Early results on fertility patterns in these studies led to a line of work on the role of deliberate delay or cessation of childbearing in producing low levels of fertility within marriage. Wang, Lee, and Campbell (1995) and Lee and Campbell (1997, pp. 83–102) showed that in the CMGPD-IL and Daoyi respectively, marital fertility was lower than in Europe, intervals between marriage and first birth and between subsequent births were much longer, and childbearing ceased much earlier. They argued that these and other patterns were consistent with deliberate behaviour to delay births and cease childbearing. These findings were the basis of the claim in Lee and Wang (1999) that contrary to the beliefs of Malthus and his successors, a fertility-based preventive check played an important role in the population dynamics of China before the 20th century, and that the rapidity of fertility decline in the 20th century in mainland China, Taiwan and Hong Kong reflected a historical legacy of adjusting fertility according to economic and other circumstances that primed the population to respond quickly when new technologies for fertility limitation became available. A vigorous debate with advocates of a Malthusian interpretation of China's historical population dynamics ensued (Campbell, Wang, & Lee, 2002; Lee, Campbell, & Wang, 2002). Campbell and Lee (2010b) revisited the issue of fertility control and showed that once heterogeneity in fecundity across couples was properly accounted for, there was clear evidence of stopping behaviour.

Paralleling our study of mortality, we moved on to map fertility differentials to illuminate influences of community, household and individual context on reproduction. Lee and Campbell (1997, pp. 133–156, pp. 177–195) compared cumulative numbers of boys born by household structure, location within the household, and socioeconomic status in Daoyi. In general, men who had privileged statuses in the household or socioeconomic hierarchy had more children. This reflected not only earlier marriage and a higher likelihood of remarriage, but in some cases higher fertility within marriage. This consistent positive relationship between privilege and reproduction is in contrast with mortality, which as noted above was in some cases the opposite of what was expected, with privileged males experiencing higher death rates. Fertility fell during times of hardship, that is when grain prices were high or when there were climatic shocks (Campbell & Lee, 2010a; Wang, Campbell, & Lee, 2010). Dong (2016) also studied the role of local family system in moderating the influence of co-resident kin on reproduction between East Asian populations.

Subsequent analyses focused primarily on fertility within marriage and such related behaviour as adoption. Wang, Lee, and Campbell (2010) revisited fertility in an expanded CMGPD-LN and demonstrated that it was linked to location in economic and household hierarchies. Campbell and Lee (2009) examined associations of marital fertility with characteristics of kin living outside the household but did not find any. Chen, Lee, and Campbell (2010) showed that fertility in Shuangcheng was positively associated with family landholding and with other measures of socioeconomic and household status. Wang and Lee (1998) showed that in the Qing imperial lineage, as many as 12.5% of sons

were adopted between related individuals, and that more generally, adoption played an important role in maintaining the continuity of the descent line and achieving other goals.

#### 4.1.2 MORTALITY

Early descriptive analysis of infant and child mortality led to an examination of female infanticide that became one of the foundations of the critique of Malthusian interpretations of China's historical population dynamics in Lee and Wang (1999). Lee, Campbell, and Tan (1992) and Lee and Campbell (1997, pp. 58–82) used indirect evidence from registered births and deaths in Daoyi to argue that families employed female infanticide or neglect to influence the number and sex composition of surviving children, and in doing so responded to economic conditions as well as their personal circumstances. As noted above, dissatisfaction with reliance on indirect evidence inspired the creation of the CMGPD-IL, which recorded the births of sons and daughters completely, as well as the deaths that occurred afterward. This led to the analysis of infant and child mortality in the Imperial Lineage in Lee, Wang, and Campbell (1994) that provided direct evidence of infanticide in the form of dramatically higher death rates for daughters in the first day and month of life, furthermore demonstrating that infanticide was not only a response to crisis or desperate poverty.

The next set of mortality studies employed event-history analysis to map patterns of mortality differentials and illuminate the influence of family, community, and institutional context on death risks. Lee and Campbell (1997, pp. 133–156, pp. 177–195) first showed that mortality rates varied according to socioeconomic status and location within the household hierarchy. Relationships were sometimes counterintuitive: male privilege was sometimes associated with higher mortality risk. Campbell and Lee (1996) showed that mortality risks depended not only on household size and composition but also on the presence or absence in the household of specific kin. Campbell and Lee (2002b) examined how household context conditioned the mortality effects of widowhood and orphanhood and showed that widows' mortality risks depended on whether they had a son. Widows who had a son were unaffected by the loss of their husband, but widows without a son experienced elevated mortality. As we describe below in our summary of findings, Hao Dong lead-authored comparisons of family contextual influences on mortality in Liaoning and Taiwan in China and northeast Japan using pooled household register data from all three locations (Dong, 2016; Dong et al., 2017).

This led to studies of the short-term consequences of economic and climatic shocks and long-term effects of public health interventions. Campbell and Lee (2000) argued based on an analysis of effects of the interactions of social status, household context and prices that there was a trade-off between privilege and mortality risk, with the mortality of privileged individuals also being more sensitive to price fluctuations. Campbell and Lee (2004) used a much larger sample from the CMGPD-LN to investigate differentials in mortality levels and mortality sensitivity to price fluctuations in more detail. Male mortality was more sensitive to grain price fluctuations than female mortality, and the response was conditioned by age, socioeconomic status, and household context. These results contributed to the comparisons between East and West in Bengtsson et al. (2004). Campbell and Lee (2010a) investigated the effects of unusually cold summers and other climatic disruptions in the years 1782–1789, 1813–1815, and 1831–1841. During the first of these periods, life expectancy fell by more than 10 years. Young males and females were especially hard hit, with the death rates of males aged 5–15 being multiplied by 8.78 and females aged 5–15 being multiplied by 4.65. Campbell (1997; 2001) assessed the effects of public health interventions in Beijing at the beginning of the 20th century and immediately after 1949 by comparing mortality in the CMGPD-IL in the 19th century to rates in Beijing at different points in time in the early, mid and late 20th century.

Recent studies investigate the consequences of family context and history for mortality later in life or in later generations. Chen et al. (2005) used the CMGPD-SC to compare the death rates of settlers in Shuangcheng according to whether their families originated in urban Beijing and its environs or from rural northeast China. She found that descendants of migrants from Beijing experienced a persistent mortality disadvantage even though state policies privileged them. Campbell and Lee (2009) used the CMGPD-LN to study how household context affected mortality in adulthood and old age. They found that men who had lost their mothers in childhood or whose mothers were 35 or older when they were born had higher death rates in adulthood, and that men experienced elevated mortality risks in old age if they were born after a short preceding birth interval, to women who were 35 or older, to a father listed as disabled, or to a father who held a salaried official position. Dong and Lee (2014) used the CMGPD-LN to examine mortality in later life of men who had migrated from one village

to another in childhood and found that they had more favourable outcomes if they had kin in their destination village. Most recently, Zang and Campbell (2018) used the CMGPD-LN to investigate how co-residence with grandparents in childhood influenced mortality in adulthood and old age.

### 4.1.3 MARRIAGE AND HOUSEHOLD

Marriage has been a fruitful area for study because marriage timing and the overall chances of marriage closely reflected household priorities and individual privilege within the household. Marriage was the direct result of an explicit decision by the household about when a son or daughter would marry, and who they would marry. By contrast, with the exception of infanticide, fertility and mortality were outcomes that were influenced by household priorities and decisions but were also subject to a variety of other unrelated influences, making associations much more difficult to interpret. High status males were more likely to marry and if widowed, remarry. The first demonstration of the positive association between male socioeconomic and household status and marriage chances was for Daoyi in Lee and Campbell (1997, pp. 133–156, pp. 177–195). In the Imperial Lineage, social status was also positively associated with male marriage chances (Lee, Wang, & Ruan, 2001). Socioeconomic status of distant but co-residing kin influenced male marriage chances, and there was also clear evidence of sequencing among unmarried males of the same generation within the household (Campbell & Lee, 2008b). Higher status females tended to marry later but almost all females married eventually (Chen, Campbell, & Lee, 2014). Remarriage chances were tied to socioeconomic status as well, with higher status widowers more likely to remarry.

We also considered other aspects of marriage, including polygyny and the effects of economic shocks. Even though polygyny was one of the most widely noted features of marriage in China before the 20th century, it was extremely rare in the rural populations covered by the CMGPD-LN and CMGPD-SC. Even in the elite Imperial Lineage, polygyny became steadily less common over time, so that by the last half of the 19th century it was rare except among close relatives of the emperor. Moreover, polygyny was used primarily to extend the reproductive span of males rather than to father children with different partners at the same time (Lee et al., 2001). In rural Liaoning, economic hardship as reflected in elevated grain prices did not have an immediate effect on marriage the way it did on mortality and fertility, but had a lagged effect because elevated female infant and children mortality disproportionately reduced the numbers of girls reaching adulthood two decades later, worsening the imbalance in the marriage market (Campbell & Lee, 2008a).

Recently we have examined assortative marriage, that is who marries whom, for insight into family preferences regarding their affinal connections. This helps delineate social, economic and institutional boundaries between groups in historical China. Our first paper on the topic examined interethnic marriage in the CMGPD-SC for insight into whether in a unique institutional setting where Han and Manchu were allowed to intermarry without being affected by rules forbidding marriage between affiliates of the Eight Banners and regular civilians, they would do so (Chen, Campbell, & Dong, 2018). We found that marriage between Manchu and Han was common and that its likelihood depended on family characteristics including a family history of intermarriage, local marriage market composition, and other factors. Our second paper on the topic examined assortative mating by education and family class label in rural Shanxi in China in the middle of the 20th century (Xing et al., 2020) and found that both were important in marriage formation, and that patterns changed little before and after 1949, when the People's Republic of China was established. This was a novel finding because while there are many studies of educational assortative marriage in China in the last half of the 20th century, there are fewer that consider the middle of the 20th century and simultaneously consider the role of class labels.

Another line of work investigates household dynamics, including the growth of households and the formation of new ones by household division. In Liaoning, a large share of the population lived in large households with many distantly related individuals living together (Lee & Campbell, 1997, pp. 105–132; Lee & Gjerde, 1986). These households were highly hierarchical, with status and privilege determined by relationship to the household head (Lee & Campbell, 1997, pp. 133–156). The head and his or her children and grandchildren of the head were most privileged, and more distant kin were less privileged. When households divided, it was typically upon the death of a head or other senior relative whose presence linked different kin groups within the household (Lee & Campbell, 1998). Household heads were predominantly male, but widows sometimes inherited the headship after the death of their husband. Household division was a liberating experience for distant relatives of the

head who previously had been at the bottom of the hierarchy but as a result of division now exercised control over the resources of the newly formed household (Campbell & Lee, 1999).

## 4.2 COMPARISONS

For twenty years from 1994 to 2014, together with Feng Wang, we analysed the CMGPD datasets as part of a collaborative international comparison: the Eurasia Project in Population and Family History. By carrying out nearly identical analyses with similar data from different settings in Europe and Asia, we compared patterns of demographic responses to economic conditions in southwestern Sweden, northeast Japan, northeast China, eastern Belgium, and northern Italy. We interacted extensively with Tommy Bengtsson, Christer Lundh and Martin Dribe who worked on Scania in southwest Sweden, Akira Hayami, Noriko Tsuya, and Satomi Kurosu who worked on Fukushima in northeast Japan, Michel Oris and George Alter who worked on eastern Belgium, and Marco Breschi, Matteo Manfredini, and Renzo Derosas who worked on northern Italy.<sup>35</sup>

The resulting comparisons of mortality (Bengtsson et al., 2004), fertility (Tsuya et al., 2010) and nuptiality (Lundh et al., 2014) revealed unanticipated similarities between East and West, the role of household context in shaping demographic behaviour, as well as unexpected differences. Key findings were that in the West, socioeconomic differences were important for shaping demographic responses to economic shocks, while in the East, socio-political differences in household context were more important. Overall, and in contrast with expectations based on Malthusian interpretations of population dynamics, demographic responses to economic shocks were weaker in the East than in the West. The emphasis on comparison by analysis of results from models that were the same across all the different datasets distinguished this effort from previous international comparisons of population and family in past times and led to novel results.

For insight into the strengths and weaknesses of the registers that were the basis of the CMGPD-LN, we also conducted comparisons of the same families recorded in the CMGPD-LN and their own genealogies. Among the materials we collected from each village we visited during our eight field trips were lineage genealogies. We transcribed these into a dataset and then compared the recording of lineage members between the CMGPD-LN and the genealogies (Campbell & Lee, 2002a). We found that as is already widely known, sons who died in infancy and childhood as well as daughters tended to be omitted from family genealogies, leading fertility estimated from genealogies to be underestimated. We also showed that fertility estimated from genealogies could be underestimated because they were more likely to omit adults who never married and married adults who did not have any surviving heirs. Whereas previous research had assumed that because of the omission of sons who died early and most daughters, fertility estimates from genealogies could be 'corrected' with an adjustment for infant and child mortality and the sex ratio at birth, the countervailing biases associated with the omission of childless adults made adjustment much more difficult, or perhaps even impossible.

We have initiated a new comparative, collaborative study of family and demographic behaviour in historical East Asia led by Hao Dong. Hao Dong harmonized datasets from northeast China, northeast Japan, Korea and Taiwan, and worked with Satomi Kurosu (Japan), and Wenshan Yang (Taiwan) on the analysis of these data. For these comparisons we have also made use of triennial Korean household registers from the county of Tansong made publicly available by a group of historians who at the time were mostly at Sungkyunkwan University and which we have turned into a longitudinal dataset through nominative linkage.<sup>36</sup> The resulting comparative studies explore how family context including presence and absence of various kin influenced demographic outcomes across East Asian populations (Dong et al., 2015a, 2015b; Dong, 2016; Dong et al., 2017).

## 4.3 SOCIAL MOBILITY, INEQUALITY AND MIGRATION

Our study of social mobility progressed from the analysis of associations in the outcomes of fathers and sons to the study of the role of networks of kin in shaping individual outcomes and finally to the study of lineages in their own right, with lineage membership as a key stratifying variable in historical

<sup>35</sup> Each of the teams had other participants with whom we interacted more sporadically.

<sup>36</sup> Cameron Campbell and Hao Dong wrote software to produce longitudinal links of records of the same individual in different registers to transform the cross-sectional data into a panel dataset (Dong et al., 2015b). The longitudinal links are available at <https://doi.org/10.14711/dataset/IVIDZV>.

Chinese society. Initial studies of father-son associations revealed that a son had a greater chance of obtaining a salaried official position in Daoyi if his father held one (Lee & Campbell, 1997, pp. 196–215). Comparison with results of studies of social mobility in 19th century North America and Europe revealed that the attainment advantages of the sons of locally elite fathers were nevertheless much less pronounced in Liaoning than in the West (Campbell & Lee, 2003). Ethnic mobility accompanied social mobility in the sense that Han men who held salaried official positions were more likely to change from Han to Manchu names (Campbell, Lee, & Elliott, 2002). In every generation, large proportions of the men in Liaoning who entered the local elite by obtaining salaried government positions not only were 'new' in the sense that not only did their father not hold a position, but neither did any of their other patrilineal relatives (Campbell & Lee, 2003). Positions held by other kin were usually a source of advantage, though not always (Campbell & Lee, 2008b).

Lineage membership was also a source of differentiation in rural Liaoning. Social and demographic outcomes, especially attainment chances and marriage chances, depended not only on individual and household characteristics but also lineage affiliation (Campbell & Lee, 2008c). There was long-term continuity not only during the Qing but between the Qing and the late 20th century in the relative status of lineages (Campbell & Lee, 2011). Socioeconomic privilege not only increased the number of children a man had, but increased the total number of descendants he had for as many as six generations, meaning that in every generation, a disproportionate share of the population was descended from the most socioeconomically privileged members of the population in previous generations (Song, Campbell, & Lee, 2015). We have also explored computational approaches to the study of lineages: Fu et al. (2018) used visualization and network techniques to study the determinants of the morphology of descent line structure.

Newer work considers inequality from an even broader perspective. Chen (2017) examines stratification based on institutional affiliation and landholding in Shuangcheng. The state specified different entitlements to land according to population category defined by institutional affiliation. These differential land entitlements affected landholding as well as access to other social and economic privileges. The residents of Shuangcheng challenged the state defined social hierarchy in some cases but at the same time reinforced it in others. Noellert (2020) examines individual-level data on Land Reform events in Shuangcheng after 1945 and found that a redistribution of power away from local strongmen paved the way for the reallocation of property, which continued to be defined by state entitlements.

We have also conducted studies of migration. The CMGPD-LN follows households when they move within Liaoning. When people leave the region entirely, usually illegally, that is recorded as well. Our first study examined the determinants of legal migration of households within Liaoning and illegal departure from the region (Campbell & Lee, 2001). Household age structure conditioned legal migration: 'younger' households with fewer elderly dependents were much more likely to migrate. Households with men who held salaried positions, meanwhile, were less likely to move. Illegal departure was more common for men who were unmarried or widowed, distant relatives of the household head, or members of smaller households. Dong et al. (2015a) compares patterns in northeast China with those in 18th and 19th century Korea and Japan.

#### 4.4 SOCIAL AND SPATIAL ORIGINS OF UNIVERSITY STUDENTS IN 20TH CENTURY CHINA

Studies based on CUSD datasets of student registration cards and other materials have illuminated shifts in the spatial and social origins of university students in China from the late 19th century to the beginning of the 21st century. Whereas during the Qing educated elites were recruited nationally via the examination system until its abolition in 1905, the educated elites who dominated Republican China in the first half of the 20th century were generally drawn from merchant and white-collar professional families in the major coastal cities (Liang et al., 2017; Ren et al., 2020). Liang et al. (2013) moreover showed that immediately after 1949, student origins at Peking University and Soochow University continued to resemble Republican-era universities, with disproportionate numbers of students coming from business and professional families in the coastal cities.

More importantly, Liang et al. (2012, 2013) also show that the introduction of standardized exams (*gaokao*) in 1955 together with a major expansion in primary and secondary education fundamentally transformed the composition of university-eligible students. In particular, the numbers of students from

farm and factory families who were the first in their families to attend college significantly increased. This pattern persisted well into the 1990s, when the share of children of professionals began to rise again. At least until 2004, however, approximately 30% of students in Peking University and 40% of students in Soochow University still originated from working-class families. This was a very different pattern from much of the West, where the students who attend the elite private universities that are the counterpart of Peking and Soochow Universities overwhelmingly come from high-income families. These findings had considerable impact on ongoing Chinese debates at the turn of the twenty-first century whether the college entrance examinations still maintained earlier opportunities for students from families of modest means, or favoured students from already well-off families.<sup>37</sup>

#### 4.5 QING OFFICIALDOM AND THE CAREERS OF OFFICIALS

Analysis of the CGED-Q has already yielded insights into Qing officialdom and the careers of officials not available from traditional approaches to the study of the Qing civil service which emphasize case studies of individuals or offices, or specific time periods. Ren et al. (2016), Chen (2019) and Chen et al. (2020) show that the central government, especially its upper reaches, were dominated by Manchu and other bannermen right up to the end of the Qing. Only a relatively small share of Han who qualified by their civil service examination performance served in the central government largely confined to the Hanlin Academy and related offices. Outside the central government, however, officials were predominantly Han, and included more holders of purchased degrees than holders of exam degrees. Median career length was just under seven years, except for bannermen and holders of *gongsheng* exam degrees, whose medium career length was three years (Chen et al., 2020). The abolition of the examination system in 1905 had little effect on the holders of exam degrees who were already officials, or on holders of exam degrees who awaited appointment. Chen, Campbell and Lee (2018) examined Banner officials at the very end of the Qing and found that their numbers and positions changed little during the New Government period, but that their share of officials declined because of an increase in the number of Han officials. Campbell (2020) shows that after the abolition of the civil service exams in 1905, men who already held exam degrees continued to be appointed at the same pace as before, and the turnover of exam degree holders who were officials was unaffected. Such results challenge claims made in other studies that the abolition of the examinations adversely affected aspiring elites (Bai & Jia, 2016).

### 5 CONCLUSION

Looking back at four decades of collaboration on the study of demographic, social and economic history, we have some reflections and observations. The first is that we have been extraordinarily fortunate in terms of finding, acquiring, and constructing the diverse, large microdata sources which are the basis of almost all our research. This has been a group effort. Our achievements in identifying new sources of microdata to understand China's past and sometimes present are increasingly due to collaborations with our colleagues in the Lee-Campbell Research Group. Notable examples are Chen Liang and Bamboo Ren's compilations of Republican-era student registration records, Yuxue Ren sharing with Cameron Campbell her ongoing work on the *jinshenlu*, Matthew Noellert and Xiangning Li's discovery of new materials on China's rural reconstruction in the mid 20th century, and recent discoveries by Bamboo Y. Ren, Yibei Wu, and Li Yang, of materials on overseas students and educated professionals.

Second, institutional support was crucial for us to acquire the microdata we discuss in this article. Institutions that very generously made materials from their holdings available include the Liaoning Provincial Archives, the First Historical Archives, the Liaoning Local History Office, the Genealogical

<sup>37</sup> An initial article publicizing these results in 2012 (Liang et al., 2012) inspired immediate and widespread public debate, summarized on our group web pages, with over 100 response articles, reposted, reprinted or rebroadcast on over 1000 Chinese media sites. These included several televised reports and discussions, most notably a widely publicized recommendation of our subsequent book, Liang et al. (2013), made by CPPCC Chairman Zhengsheng Yu during a 2014 meeting on education held in conjunction with the ongoing Chinese People's Political Consultative Congress. More recently, an extended video commentary on the book by Yu Liang posted on the Chinese website Bilibili on July 12, 2020 had been viewed 240,000 times and liked 22,000 times by July 15, 2020.

Society of Utah (now called [FamilySearch](#)), the Shanxi University Research Centre for Chinese Social History, the Shuangcheng County Archives and numerous universities in China and the United States, such as Peking University, Soochow University, Shanghai Jiao Tong University, Tsinghua University, Zhejiang University, and Columbia University. For other datasets, most notably the CGED-Q, we relied on materials that had been published or made publicly available for download. The most important of these is the Tsinghua University Library collection of *jinshenlu* published as *Qingdai Jinshenlu Jicheng* that along with *jinshenlu* editions made available by the Harvard Yenching Library and Columbia University Library were the basis of most of the CGED-Q.

Third, we also benefitted from many other individuals who contributed to our construction of these datasets. Listing everyone here would be impossible, but we can single out some individuals who played especially important roles. Ju Deyuan, Robert Eng, Alice Suen, Anna Chi and others helped James Lee initiate research on Daoyi. Mel Thatcher arranged for access to the collections of the Genealogical Society of Utah. Ts'ui-jung Liu, Sufen Liu, and Huimin Lai at the Academia Sinica facilitated the creation of the CMGPD-IL and one of the CMGPD-LN register series. Among the coders at the Academia Sinica, Shu-mei Tsay made the most contributions to the CMGPD-IL and CMGPD-LN. Shuang Chen, Matt Noellert, and Bijia Chen coordinated and oversaw the data entry of the CMGPD-SC, CRRD-LR and CRRD-SQ, and CGED-Q, respectively. Similarly, Liang Chen, Bamboo Ren, Hao Zhang, Yibei Wu, and Li Yang initiated, coordinated, and oversaw the creation of various subsets of the CUSD and CPOD. Hao Dong helped create longitudinal links for the Korean registers and led the efforts to harmonize the CMGPD and other datasets from Japan, Korea, and Taiwan. Many, many coders worked tirelessly to enter all these data. It is not possible to list all of them here, but we highlight six who made especially large contributions over extended periods of time: Huicheng Sun, Jiyang, and Xing Xiao entered much of the CMGPD-LN, CMGPD-SC, CRRD-LR, and together with Xiaodong Ge, Yibei Liu, and Mi Zhao, the CGED-Q.

Fourth, we could not have proceeded without generous institutional and occasionally personal support. James Lee began his career at the California Institute of Technology and Cameron Campbell met him there while an undergraduate. In retrospect, Caltech was one of only a few places that in the early 1980s would support an assistant and then associate and finally full professor of humanities/history to carry out quantitative research on China. It is also hard to imagine that at any other institution, a sophomore studying electrical engineering who had a side interest in Chinese history but no language ability could walk in to a history professor's office and after some discussions outline a plan to reorganize data management and analysis for an ongoing project and then become a collaborator. In graduate school at Penn and then as an assistant, associate and then full professor and sociology at UCLA, Campbell was supported by mentors and then colleagues even though his work was esoteric.

Internal funding and administrative support from the California Institute of Technology, the University of Michigan, UCLA, Peking University, the Hong Kong University of Science and Technology, Shanghai Jiao Tong University, and from other universities in China allowed us to greatly expand our data acquisition and dataset construction, as well as to apply for sustained research support from the National Institutes of Health in the USA, the National Science Council in Taiwan, the National Natural Science Foundation in mainland China, and the Research Grants Council in Hong Kong. Equally important, these universities provided opportunities and sometimes funding to collaborate with the graduate students, postdoctoral fellows, and visiting professors who made crucial contributions to, and in some cases led, the data construction and research which define much of our work over the last forty years and hopefully for many years in the future. We are especially grateful to Myron Guttman, who as head of the Inter-university Consortium for Political and Social Research, provided guidance and support when we were scaling up our operations and seeking for the first time substantial extramural funding.

Long-term collaborations played a key role in advancing our work. The most sustained and for us influential collaboration was the twenty years we spent working with colleagues from a variety of countries and disciplines on the Eurasia Project in Population and Family History. Interactions with project participants stimulated us to broaden our range of research topics, learn and apply more advanced methods, and seek opportunities for comparisons for our other projects. The camaraderie that grew out of frequent, sustained interaction with others working with data like ours and on similar topics was also important for our own morale. We have especially fond memories of our fruitful two decade-long collaboration with Feng Wang which produced a number of discrete studies as well as Lee and Wang (1999) and Tsuya et al. (2010) and our collaborations with Tommy Bengtsson and Noriko Tsuya, which involved reciprocal visits by us in Lund and Tokyo and by them in Pasadena.

Many short-term collaborations on specific papers and long-term projects were also important. Yizhuang Ding and Songyi Guo advised on the CMGPD-LN and with the help of Gao Jing from the Liaoning Provincial Gazetteer Office we conducted fieldwork with them that resulted in Ding et al. (2004). Songyi Guo also shared his expertise with us when we were constructing and analysing the CMGPD-IL. We have been fortunate to co-author with a variety of others on papers or sets of papers using our datasets, including Lawrence Anthony, Mark Elliott, Robert Eng, William Lavelly, Chris Myers, Xi Song, Alice Suen, Guofu Tan, Emma Zang, and Siwei Fu and other members of Huamin Qu's group. Similarly, we have benefited from interaction with Akira Hayami, Kuentae Kim, Satomi Kurosu, Sangkuk Lee, Ts'ui-jung Liu, Byun-giu Son, Noriko Tsuya, Wenshan Yang, and other collaborators.

Looking back, we believe that a distinctive feature of our research that has been central to our success has been our inductive, data-driven approach which emphasizes discovery of facts about demographic behaviour and family, social and economic organization through empirical analysis of the datasets we have constructed. We have always started with data that we thought might help us investigate a topic of general interest, and then through exploratory and descriptive analysis sought to uncover key patterns of demographic behaviour and family and social organization, moving to carefully specified regression-based models only after extensive work to verify the data and then elaborate on relationships and patterns discovered in descriptive analysis. While this approach is time-consuming, sometimes taking years to locate, access, enter and clean data for the analyses that led to our major results, we believe that the result has been a fundamental transformation in our understanding of basic patterns of family, social and economic organization in China in the past and increasingly up to the very recent present.

Our strong belief in the importance of a scholarship of microdata-driven empirical discovery as opposed to a scholarship of interpretation is because there is much about the Chinese past we do not know, or worse, think we know, but are wrong.<sup>38</sup> Whenever feasible and permissible, therefore, we have coded datasets in their entirety and devoted considerable time and energy to produce detailed documentation and User Guides to accompany possible future public data releases, and create a complete, permanent resource to be used by ourselves and others to study a wide range of topics. Related projects described in this special issue are underway for a variety of other locations around the world and we look forward to a future where such datasets are produced and used routinely in social science and in history to discover basic facts about life in the past, right up to the time a few decades ago when longitudinal surveys and other sources become available.

## ACKNOWLEDGEMENTS

While working on this manuscript, Lee and Campbell received support from HK RGC GRF 16602117 (Lee PI), HK RGC GRF 16601718 (Campbell PI), and HK RGC GRF 16600017 (Campbell PI). We are grateful to the members of the Lee-Campbell Group for their comments, corrections, and other feedback.

## REFERENCES

- Bai, Y., & Jia, R. (2016). Elite recruitment and political stability: The impact of the abolition of China's civil service exam system. *Econometrica*, 84(2), 677–733. doi: [10.3982/ECTA13448](https://doi.org/10.3982/ECTA13448)
- Bengtsson, T., Campbell, C. D., Lee, J. Z. et al. (2004). *Life under pressure: Mortality and living standards in Europe and Asia, 1700–1900*. Cambridge, MA: MIT Press.

38 A good example world-wide would be Malthus' previously commonly accepted description of the Chinese demographic system and its consequences (Lee & Wang, 1999). Another largely Chinese example would be the debate over opportunity for first-college students from families of marginal means to attend some of China's most competitive universities. See too our acknowledgement in Bengtsson et al. (2004) of our own past misunderstandings of mortality and related behavior and our description of the processes by which we were able to correct ourselves and achieve new understandings of the past (435–439).

- Cai, S., Lee, J. Campbell, C. D. & Myers, C. (1994). Zongrenfu dang'an diannaoku de jianli, fenxi, liyong, jiqi kunnan [Machine analysis and data coding of the Qing imperial lineage]. In J. Lee & S. Guo (Eds.), *Qingdai huangzu renkou xingwei he shehui huanjing [The Qing imperial lineage: Social structure and population behavior]* (pp. 191–204). Beijing: Peking University Press.
- Campbell, C. D. (1997). Public health efforts in China before 1949 and their effects on mortality: The case of Beijing. *Social Science History*, 21(2), 179–218. doi: [10.2307/1171273](https://doi.org/10.2307/1171273)
- Campbell, C. D. (2001). Mortality change and the epidemiological transition in Beijing, 1644–1990. In T.-j. Liu, J. Lee, D. S. Reher, O. Saito, & F. Wang (Eds.), *Asian population history* (pp. 221–247). Oxford: Oxford University Press.
- Campbell, C. D. (2016). Large-scale collaboration and comparison in historical demography: Reflections on the Eurasia project. In K. Matthijs, S. Hin, J. Kok, & H. Matsuo (Eds.), *The future of historical demography: Upside and inside out* (pp. 193–196). Leuven: Acco.
- Campbell, C. D. (2020). Qingmo keju tingfei dui shiren wenguan qunti de yingxiang-jiyu weiguan dashuju de hongguan xin shijiao [The influence of the abolition of the examinations at the end of the Qing on the holders of exam degrees]. *Shehui kexue jikan [Social Science Journal]*, 4(249), 156–166.
- Campbell, C. D., Dong, H., & Lee, J. Z. (2016). *China multigenerational panel datasets training guide*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.
- Campbell, C. D., Wang, F., & Lee, J. Z. (2002). Pretransitional fertility in China. *Population and Development Review*, 28(4), 735–750. doi: [10.1111/j.1728-4457.2002.00735.x](https://doi.org/10.1111/j.1728-4457.2002.00735.x)
- Campbell, C. D., & Lee, J. Z. (1996). A death in the family: Household structure and mortality in rural Liaoning: Life-event and time-series analysis, 1792–1867. *The History of the Family*, 1(3), 297–328. doi: [10.1016/S1081-602X\(96\)90026-3](https://doi.org/10.1016/S1081-602X(96)90026-3)
- Campbell, C. D., & Lee, J. Z. (1999). Causes and consequences of household division in northeast China, 1789–1909. In Y. Ding, S. Guo, & J. Lee (Eds.), *Hunyin yu jiating: Dongxi bijiao shi [Marriage and family, East-West comparative methods]* (pp. 1–32). Beijing: Peking University Press.
- Campbell, C. D., & Lee, J. Z. (2000). Price fluctuations, family structure and mortality in two rural Chinese populations: Household responses to economic stress in eighteenth- and nineteenth-century Liaoning. In T. Bengtsson & O. Saito (Eds.), *Population and the economy: From hunger to modern economic growth* (pp. 371–420). Oxford: Oxford University Press.
- Campbell, C. D., & Lee, J. Z. (2001). Free and unfree labour in Qing China: Emigration and escape among the bannermen of northeast China, 1789–1909. *The History of the Family*, 6(4), 455–476. doi: [10.1016/S1081-602X\(01\)00088-4](https://doi.org/10.1016/S1081-602X(01)00088-4)
- Campbell, C. D., & Lee, J. Z. (2002a, publ. 2006). State views and local views of population: Linking and comparing genealogies and household registers in Liaoning, 1749–1909. *History and Computing*, 14(1-2), 9–29.
- Campbell, C. D., & Lee, J. Z. (2002b). When husbands and parents die: Widowhood and orphanhood in late imperial Liaoning, 1789–1909. In R. Derosas & M. Oris (Eds.), *When dad dies* (pp. 313–334). Bern: Peter Lang.
- Campbell, C. D., & Lee, J. Z. (2003). Social mobility from a kinship perspective: Rural Liaoning, 1789–1909. *International Review of Social History*, 48(1), 1–26. doi: [10.1017/S0020859002000901](https://doi.org/10.1017/S0020859002000901)
- Campbell, C. D., & Lee, J. Z. (2004). Mortality and household in seven Liaodong populations. In T. Bengtsson, C. D. Campbell, J. Z. Lee et al., *Life under pressure: Mortality and living standards in Europe and Asia, 1700–1900* (293–324). Cambridge, MA: MIT Press.
- Campbell, C. D., & Lee, J. Z. (2008a). Economic conditions and male first marriage in northeast China, 1749–1909. *Sungkyun Journal of East Asian Studies*, 8(1), 17–42. doi: [10.21866/esjeas.2008.8.1.002](https://doi.org/10.21866/esjeas.2008.8.1.002)
- Campbell, C. D., & Lee, J. Z. (2008b). Kin networks, marriage, and social mobility in late imperial China. *Social Science History*, 32(2), 175–214. doi: [doi.org/10.1017/S0145553200010749](https://doi.org/10.1017/S0145553200010749)
- Campbell, C. D., & Lee, J. Z. (2008c). Villages, descent groups, households and individual outcomes in rural Liaoning, 1789–1909. In T. Bengtsson & G. Mineau (Eds.), *Kinship and demographic behaviour in the past* (pp. 73–104). Springer.
- Campbell, C. D., & Lee, J. Z. (2009). Long-term mortality consequences of childhood family context in Liaoning, China, 1749–1909. *Social Science and Medicine*, 68(9), 1641–1648. doi: [10.1016/j.socscimed.2009.02.017](https://doi.org/10.1016/j.socscimed.2009.02.017)
- Campbell, C. D., & Lee, J. Z. (2010a). Demographic impacts of climatic fluctuations in northeast China, 1749–1909. In S. Kurosu, T. Bengtsson, & C. D. Campbell (Eds.), *Demographic responses to economic and environmental crisis* (pp. 107–132). Kashiwa: Reitaku University Press.

- Campbell, C. D., & Lee, J. Z. (2010b). Fertility control in historical China revisited: New methods for an old debate. *History of the Family*, 15(4), 370–385. doi: [10.1016/j.hisfam.2010.09.003](https://doi.org/10.1016/j.hisfam.2010.09.003)
- Campbell, C. D., & Lee, J. Z. (2011). Kinship and the long-term persistence of inequality in Liaoning, China, 1749–2005. *Chinese Sociological Review*, 44(1), 71–104. doi: [10.2753/CSA2162-0555440105](https://doi.org/10.2753/CSA2162-0555440105)
- Campbell, C. D., Lee, J. Z., & Elliott, M. (2002). Identity construction and reconstruction: Naming and Manchu ethnicity in northeast China, 1749–1909. *Historical Methods*, 35(3), 101–116. doi: [10.1080/01615440209601201](https://doi.org/10.1080/01615440209601201)
- Chen, B. (2019). *Origins and career patterns of the Qing government officials (1850–1912): Evidence from the China Government Employee Dataset-Qing (CGED-Q)* (PhD dissertation). Hong Kong University of Science and Technology Division of Social Science, China
- Chen, B., Campbell, C. D., & Dong, H. (2018). Interethnic marriage in northeast China, 1866–1913. *Demographic Research*, 38(34), 929–966. doi: [10.4054/DemRes.2018.38.34](https://doi.org/10.4054/DemRes.2018.38.34)
- Chen, B., Campbell, C. D., & Lee, J. Z. (2018). Qingmo xingzheng qianhou qiren yu zongshi guanyuan de guanzhi bianhua chutan-yi jinshenlu shujuku wei cailiao de fenxi [The transition of banner and imperial lineage officials during the late Qing reform period: Evidence from the Qing Jinshenlu Database]. *Qingshi Yanjiu [Studies in Qing History]*, 2018(4), 10–20.
- Chen, B., Campbell, C. D., Ren, Y., & Lee, J. Z. (2020). Big data for the study of Qing officialdom: The China government employee database-Qing (CGED-Q). *The Journal of Chinese History*, 4(2):431–460. doi: [0.1017/jch.2020.15](https://doi.org/0.1017/jch.2020.15)
- Chen, S. (2009). *Where urban migrants meet rural settlers: State categories, social boundaries, and wealth stratification in northeast China, 1815–1913* (PhD Dissertation). University of Michigan, Department of History, USA.
- Chen, S. (2017). *State-sponsored inequality. The banner system and social stratification in northeast China*. Stanford: Stanford University Press.
- Chen, S., Campbell, C. D., & Lee, J. Z. (2005, publ. 2006). Vulnerability and resettlement: Mortality differences in northeast China by place of origin, 1870–1912, Comparing urban and rural migrants. *Annales de Démographie Historique*, 2005(2), 47–79.
- Chen, S., Campbell, C. D., & Lee, J. Z. (2014). Categorical inequality and gender difference: Marriage and remarriage in northeast China, 1749–1913. In C. Lundh, S. Kurosu et al. (Eds.), *Similarity in difference: Marriage in Europe and Asia, 1700–1900* (pp. 393–438). Cambridge, MA: MIT Press
- Chen, S., Lee, J. Z., & Campbell, C. D. (2010). Wealth stratification and reproduction in northeast China, 1866–1907. *History of the Family*, 15(4), 386–412. doi: [10.1016/j.hisfam.2010.10.001](https://doi.org/10.1016/j.hisfam.2010.10.001)
- Ding, Y., Guo, S., Lee, J. Z., & Campbell, C. D. (2004). *Liaodong yimin zhong de qiren shehui [Banner society and the settlement of eastern Liaodong]*. Shanghai: Shanghai shehui kexue chubanshe.
- Dong, H. (2016). *Patriarchy, family system and kin effects on individual demographic behaviour throughout the life course: East Asia, 1678–1945* (PhD dissertation). Hong Kong University of Science and Technology Division of Social Science, China.
- Dong, H., Campbell, C. D., Kurosu, S., & Lee, J. Z. (2015a). Household context and individual departure: The case of escape in three 'unfree' East Asian populations, 1700–1900. *Chinese Journal of Sociology*, 1(4), 513–539. doi: [10.1177/2057150X15614547](https://doi.org/10.1177/2057150X15614547)
- Dong, H., Campbell, C. D., Kurosu, S., Yang, W., & Lee, J. Z. (2015b). New sources for comparative social science: Historical population panel data from East Asia. *Demography*, 52(3), 1061–1088. doi: [10.1007/s13524-015-0397-y](https://doi.org/10.1007/s13524-015-0397-y)
- Dong, H., & Lee, J. Z. (2014). Kinship matters: Long-term mortality consequences of childhood migration, historical evidence from northeast China, 1792–1909. *Social Science and Medicine*, 119, 274–283. doi: [10.1016/j.socscimed.2013.10.024](https://doi.org/10.1016/j.socscimed.2013.10.024)
- Dong, H., Manfredini, M., Kurosu, S., Yang, W., & Lee, J. Z. (2017). Kin and birth order effects on male child mortality: Three East Asian populations, 1716–1945. *Evolution and Human Behaviour*, 38(2), 208–216. doi: [10.1016/j.evolhumbehav.2016.10.001](https://doi.org/10.1016/j.evolhumbehav.2016.10.001)
- Fu, S., Dong, H., Cui, W., Zhao, J., & Qu, H. (2018). How do ancestral traits shape family trees over generations? *IEEE Transactions on Visualization and Computer Graphics*, 24(1), 205–214. doi: [10.1109/TVCG.2017.2744080](https://doi.org/10.1109/TVCG.2017.2744080)
- Gautier, E., & Henry, L. (1958). *La population de Crulai, paroisse normande: Étude historique*. Travaux et documents – Institut national d'études démographiques, vol. 33. Paris: Presses universitaires de France.
- Ho, P.-t. (1959). *Studies on the Population of China, 1368–1953*. Harvard East Asian Series 4. Cambridge, MA: Harvard University Press.

- Ju, D. (1994). Qingdai huangzu renkou ciji [Demographic records of the Qing imperial lineage]. In J. Z. Lee & S. Guo (Eds.), *Qingdai huangzu renkou xingwei he shehui huanjing [The Qing imperial lineage: Social structure and population behaviour]* (pp. 170–190). Beijing: Peking University Press.
- Laslett, P., & Wall, R. (1972). *Household and family in past time: Comparative studies in the size and structure of the domestic group over the last three centuries in England, France, Serbia, Japan, and colonial North America, with further materials from western Europe*. New York: Cambridge University Press.
- Lavelly, W., Lee, J. Z., & Wang, F. (1990). Chinese demography: The state of the field. *Journal of Asian Studies*, 50(1), 807–834. doi: <https://doi.org/10.2307/2058237>
- Lee, J. Z., Anthony, L., & Suen, A. (1988). Liaoning sheng chengren siwang lu, 1796–1819 [Adult mortality in Liaoning, 1796–1819]. In *Qingzhu diyi lishi dang'an guan liushi zhounian lunwen ji [Proceedings of the Symposium on the Occasion of the Sixtieth Anniversary of the First Historical Archives]*, Volume 2 (pp. 885–898). Beijing: Zhonghua Shuju.
- Lee, J. Z., & Eng, R. (1984). Population and family history in eighteenth-century Manchuria: Preliminary results from Daoyi 1774–1798. *Ch'ing-shih wen-t'i*, 5(1), 1–55.
- Lee, J. Z., & Campbell, C. D. (1997). *Fate and fortune in rural china: Social organization and population behaviour in rural Liaoning, 1774–1873*. Cambridge: Cambridge University Press.
- Lee, J. Z., & Campbell, C. D. (1998). Headship succession and household division in three Chinese banner serf populations, 1789–1909. *Continuity and Change*, 13(1), 117–141. doi: [10.1017/S0268416098003063](https://doi.org/10.1017/S0268416098003063)
- Lee, J. Z., Campbell, C. D., & Anthony, L. (1995). A century of mortality in rural Liaoning, 1774–1873. In S. Harrell (Ed.), *Chinese Historical Microdemography* (pp. 163–182). Berkeley: University of California Press.
- Lee, J. Z., Campbell, C. D., & Chen, S. (2010). *China multi-generational panel dataset, Liaoning (CMGPD-LN) 1749–1909. User guide*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.
- Lee, J. Z., Campbell, C. D., & Wang, F. (1993). An introduction to the demography of the Qing imperial lineage, 1644–1911. In R. Schofield & D. Reher (Eds.), *Old and new methods in historical demography* (pp. 361–382). Oxford: Oxford University Press.
- Lee, J. Z., Campbell, C. D., & Wang, F. (2002). Positive checks or Chinese checks? *Journal of Asian Studies*, 61(2), 591–607. doi: [10.2307/2700301](https://doi.org/10.2307/2700301)
- Lee, J. Z., Campbell, C. D., & Tan, G. (1992). Infanticide and family planning in rural Liaoning, 1774–1873. In L. Li & T. Rawski (Eds.), *Chinese History in Economic Perspective* (pp. 149–176). Berkeley: University of California Press.
- Lee, J. Z., & Wang, F. (1999). *One quarter of humanity: Malthusian mythology and Chinese realities, 1700–2000*. Cambridge, Mass.: Harvard University Press.
- Lee, J. Z., Wang, F., & Campbell, C. D. (1994). Infant and child mortality among the Qing nobility: Implications for two types of positive check. *Population Studies*, 48(3), 395–412. doi: [10.1080/0032472031000147936](https://doi.org/10.1080/0032472031000147936)
- Lee, J. Z., Wang, F., & Ruan, D. (2001). Nuptiality among the Qing nobility: 1600–1900. In T.-j. Liu, J. Z. Lee, D. Reher, O. Saito, & F. Wang (Eds.), *Asian Historical Demography* (pp. 353–373). Oxford: International Studies in Demography, Oxford University Press.
- Lee, J. Z., & Gjerde, J. (1986). Comparative household morphology of stem, joint, and nuclear household systems: Norway, China, and the United States. *Continuity and Change*, 1(1), 89–111. doi: [10.1017/S0268416000000084](https://doi.org/10.1017/S0268416000000084)
- Lee, J. Z., & Guo, S. (Eds.). (1994). *Qingdai huangzu renkou xingwei he shehui huanjing [The Qing imperial lineage: Social structure and population behaviour]*. Beijing: Peking University Press.
- Lee, R., & Steckel, R. H. (2006). Life under pressure: An appreciation and appraisal. *Historical Methods*, 39(4), 171–176. Retrieved from [https://u.demog.berkeley.edu/~rlee/oldceda/Publications/pdfs/rlee/LeeSteckel\\_LifeUnderPressureReview\\_HistoricalMethods2006.pdf](https://u.demog.berkeley.edu/~rlee/oldceda/Publications/pdfs/rlee/LeeSteckel_LifeUnderPressureReview_HistoricalMethods2006.pdf)
- Liang, C., Dong, H., Ren, Y., & Lee, J. Z. (2017). Jiangshan dai you rencai chu, geling fengsao shushinian [Changes in the social and geographical origins of China's educated elites, 1865–2014], *Shehuixue yanjiu [Sociology Research]*, 2017(3), 48–70.
- Liang, C., Lee, J. Z., Zhang, H., Li, L., Ruan, D., Campbell, C. D., & Yang, S. (2012). Wusheng de geming: Beijing daxue yu suzhou daxue xuesheng shehui laiyan yanjiu [Silent revolution: The social origins of Peking University and Soochow University undergraduates, 1952–2002]. *Zhongguo shehui kexue [Chinese Social Science]*, 2012(1), 99–119.

- Liang, C., Zhang, H., Li, L., Ruan, D., Campbell, C. D., & Lee, J. Z. (2013). *Wusheng de geming: Beijing daxue, Suzhou daxue xuesheng shehui lai yuan yanjiu, 1949–2002* [Silent revolution: Social origins of Peking university and Suzhou university students, 1949–2002]. Beijing: Sanlian Books.
- Liu, T.-j., Lee, J. Z., Reher, D., Saito, O., & Wang, F. (Eds.) (2001). *Asian Population History*. Oxford: International Studies in Demography, Oxford University Press.
- Lundh, C., Kurosu, S. et al. (2014). *Similarity in difference: Marriage in Europe and Asia, 1700–1900*. Cambridge, MA: MIT Press.
- Noellert, M. (2020). *Power over property: The political economy of communist land reform in China*. Ann Arbor: University of Michigan Press.
- Ren, B., Liang, C., & Lee, J. Z. (2020). Mutable inequalities: Meritocracy and the making of the Chinese Academe 1912–1952, A data analytic approach. *China Quarterly*. Accepted for publication.
- Ren, Y., Chen, B., Hao, X., Campbell, C. D., & Lee, J. Z. (2016). Qingdai jinshenlu lianghua shujuku yu guangliao qunti yanjiu [The Qing jinshenlu database: A new source for the study of Qing officials]. *Qingshi Yanjiu* [Qing History Research], 2016(4), 61–77.
- Ren, Y., Chen, B., Hao, X., Campbell, C. D., & Lee, J. Z. (2019). *Zhongguo lishi guanyuan lianghua shujuku-qingdai jinshenlu 1900–1912 shidian gongkaiban yonghu zhinan* [The China government employee database-Qing Jinshenlu 1900–1912 Public Release User Guide]. Hong Kong: The Hong Kong University of Science and Technology.
- Song, X., Campbell, C. D., & Lee, J. Z. (2015). Ancestry matters: Patrilineage growth and extinction. *American Sociological Review*, 80(3), 574–602. doi: [10.1177/0003122415576516](https://doi.org/10.1177/0003122415576516)
- Tsuya, N.O., Wang, F., Alter, G., Lee, J. Z. et al. (2010). *Prudence and pressure: Reproduction and human agency in Europe and Asia, 1700–1900*. Cambridge, MA: MIT Press.
- Wang, F., Campbell, C. D., & Lee, J. Z. (2010). Agency, hierarchies, and reproduction in northeastern China, 1789 to 1840. In N. Tsuya, F. Wang, G. Alter, J. Z. Lee et al., *Prudence and pressure: Reproduction and human agency in Europe and Asia, 1700–1900* (pp. 287–316). MIT Press.
- Wang, F., & Lee, J. Z. (1998). Adoption among the Qing nobility and its implications for Chinese demographic behaviour. *History of the Family*, 3(3), 411–428. doi: [10.1016/S1081-602X\(99\)80255-3](https://doi.org/10.1016/S1081-602X(99)80255-3)
- Wang, F., Lee, J. Z., & Campbell, C. D. (1995). Marital fertility control among the Qing nobility: Implications for two types of preventive check. *Population Studies*, 49(3), 383–400. doi: [10.1080/0032472031000148736](https://doi.org/10.1080/0032472031000148736)
- Wang, H., Chen, S., Dong, H., Noellert, M., Campbell, C. D., & Lee, J. Z. (2013). *China multi-generational panel dataset, Shuangcheng (CMGPD-SC) 1866–1914. User Guide*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.
- Wang, L. (2012). *Qingdai huangzu chengyuan de xianfu shenfen yu guanzhi huode, 1644–1911* [Ascribed status and official attainment by members of the Qing Imperial Lineage, 1644–1911] (PhD Dissertation). Peking University Department of Sociology, China
- Wang, L., Lee, J. Z., & Campbell, C. D. (2010). Institutions and inequality: Comparing the Zongshi and the Jueluo in the Qing imperial lineage. *Sungkyun Journal of East Asian Studies*, 10(1), 33–61. doi: [10.21866/esjeas.2010.10.1.003](https://doi.org/10.21866/esjeas.2010.10.1.003)
- Wolf, A., & Huang, C.-s. (1980). *Marriage and adoption in China, 1895–1945*. Stanford: Stanford University Press.
- Xing, L., Campbell, C. D., Li, X., Noellert, M., & Lee, J. Z. (2020). Education, class and assortative marriage in rural Shanxi, China in the mid-twentieth century. *Research in Social Stratification and Mobility*, 66(2020), 100460, 1–15. doi: [10.1016/j.rssm.2019.100460](https://doi.org/10.1016/j.rssm.2019.100460)
- Zang, E., & Campbell, C. D. (2018). Males' later life mortality consequences of coresidence with paternal grandparents: Evidence from northeast China, 1789–1909. *Demography*, 55(2), 435–457. doi: [10.1007/s13524-018-0653-z](https://doi.org/10.1007/s13524-018-0653-z)

HISTORICAL LIFE COURSE STUDIES  
VOLUME 9 (2020), published 03-12-2020

## A Database for the Future

### Major Contributions from 47 Years of Database Development and Research at the Demographic Data Base

Sören Edvinsson

CEDAR, Centre for Demographic and Ageing Research, Umeå University, Sweden

Elisabeth Engberg

CEDAR, Centre for Demographic and Ageing Research, Umeå University, Sweden

#### ABSTRACT

The Demographic Data Base (DDB) at the Centre for Demographic and Ageing Research (CEDAR) at Umeå University has since the 1970s been building longitudinal population databases and disseminating data for research. The databases were built to serve as national research infrastructures, useful for addressing an indefinite number of research questions within a broad range of scientific fields, and open to all academic researchers who wanted to use the data. A countless number of customised datasets have been prepared and distributed to researchers in Sweden and abroad and to date, the research has resulted in more than a thousand published scientific reports, books, and articles within a broad range of academic fields. While there has long been a clear predominance of research within the humanities and social sciences, it has always been used for research in other fields as well, for example medicine. In this article, we first give a brief presentation of the DDB and its history, characteristics, and development from the 1970s to the present. It includes an overview of the research based on the DDB databases, with a focus on the databases POPUM and POPLINK with individual-level data. A number of major traits of the research from 1973 to now have been outlined, showing the breadth of the research and highlighting some major contributions, with a focus on work that would have been very difficult to perform without data from the DDB.

**Keywords:** Historical databases, Life courses, Population studies, Demography, Sociology, History, Life sciences

DOI article: <https://doi.org/10.51964/hlcs9305>

© 2020, Edvinsson, Engberg

This open-access work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/), which permits use, reproduction & distribution in any medium for non-commercial purposes, provided the original author(s) and source are given credit. See <http://creativecommons.org/licenses/>.

## 1 INTRODUCTION

The Demographic Data Base (DDB) at the Centre for Demographic and Ageing Research (CEDAR) at Umeå University has been committed since the 1970s to building longitudinal population databases and disseminating data for research. In contrast to many other similar ventures, already from the start the databases at the DDB were built to serve as research infrastructures, useful for addressing an indefinite number of research questions within a broad range of scientific fields, and open to all academic researchers who wanted to use the data. The DDB's role as a national research infrastructure was defined a few years later in the government ordinance of 1978 (UHÄ-FS, 1978, 75; 1990, 8), and over the years this privileged position has been both a strength of and a special challenge to the DDB. Since its start in 1973, a countless number of customised datasets have been produced and distributed through specified contracts to researchers in Sweden and abroad. To date, the research has resulted in more than a thousand published scientific reports, books, and articles within a broad range of academic fields, and has been used for almost 70 dissertations (DDB, 2016). Data has been retrieved from the DDB databases by academic researchers in Sweden, Europe, North America, Asia and Australia and has been used for research in a large variety of academic fields. Sociology, statistics, demography, geography, history, economics, medicine and cultural studies, are just some examples. Today, the DDB owns and administers three main research databases: POPUM, with individual-level data from Swedish parishes covering the period 1680–1900; POPLINK, with similar data but over a longer time span, until around 1950; and TABVERK, with aggregate statistics from all Swedish parishes for the period 1749–1859.

In this article, we first give a brief presentation of the DDB and its history, characteristics, and development from the 1970s to the present. The second part includes an overview of the research based on the DDB databases, with a focus on the databases POPUM and POPLINK with individual-level data. Considering that we are talking about more than a thousand published studies, it would be too great an endeavour to give a full and fair presentation of all the research. Instead, we have outlined a number of major traits of the research from 1973 to now, thereby showing the breadth of the research and highlighting some major contributions, with a focus on work that would have been very difficult to perform without data from the DDB.

## 2 FROM PARISH RECORDS TO LIFE-COURSE DATA

### 2.1 A VISIONARY ENTERPRISE

While the DDB was first established in 1973, its inception can be traced back to the pioneering and innovative work of Professor Egil Johansson (1933–2012) in the mid-1960s, studying the history of literacy in Sweden with quantitative methods and information from the longitudinal Swedish parish registers (Johansson, 1969/70). Realising the potential of these sources, he also managed to convince others of the value of digitising them. Although this was an extensive, costly, and visionary enterprise, undertaken at a time long before personal computers and modern software had been invented, the proposal aroused great interest within the scientific community (Brändström, 2009). This can be attributed to the increasing focus on quantitative social and economic history research in Europe at the time. With Louis Henry, France had witnessed the birth of historical demography, and since the 1950s parish registers had been used in pioneering studies of population and fertility (Rosenthal, 2003). In the mid-1960s, around the same time as Egil Johansson embarked on his project, Peter Laslett published his seminal work *The World we Have Lost* (Laslett, 1965), which with similar sources and quantitative methods questioned and overturned popular conceptions about life in past Britain. Johansson's work on literacy showed that the Swedish sources, combined with modern information technology, had no less potential to make valuable contributions to social and demographic history, to the history of ordinary people, and to our knowledge of life and death among populations of the past.

The first steps of database-building were taken in 1973 through a government-funded employment project under the Swedish National Archives, creating jobs in parts of northern Sweden where opportunities for those providing labour, particularly female labour, were scarce. The new enterprise was a labour-intensive project, and this form of funding created a win-win situation for all parties involved. Between 1973 and 1982, data entry units were established in six different locations in northern Sweden, funded by provisional contributions from the National Labour Market Board. After five years, in 1978, the Swedish National Archives stepped down as the responsible authority and Umeå University assumed the future organisational and financial

responsibility for DDB. The future development required more sustainable and long-term funding than the short-term employment policy-related measures could provide, and formally becoming a part of the university organisation was considered a good solution. The transition from the Swedish National Archives to Umeå University in no way meant that the national interest in the project was diminished, however. The governmental hand was still there, in the form of earmarked funding in the annual university budget and in the remaining governmental ordinance defining the DDB's national role and mission, issued the same year (Brändström, 2009). Today, the DDB receives its core funding from Umeå University, supplemented by funding from external project grants.

As mentioned, data was not collected to serve any particular research project or group, or with certain research questions in mind. The intention was instead that the databases would be long-lasting resources for the future, usable for all sorts of research questions within all kinds of scientific fields, nationally as well as internationally. To be able to achieve this, it was very important to find ways to listen to the needs of the research community. Database-building is a long-term commitment and a costly enterprise that requires a combination of constantly having your ear to the ground and wise strategic planning. One important source of input from the research community has been active participation in scientific meetings and conferences. From the 1970s onward, a group of researchers was formed around the DDB, pursuing research on the data and developing research methodology. The idea was also that, through their networks, they would inspire and initiate national and international research on the DDB data. A further development came in 1990 when the Centre for Population Studies was established, with the aim to promote local, national, and international population research based on DDB data, and to develop international networks. Organising international conferences and workshops, around the data or related to different research topics in historical demography, has also been a vital part of the strategy to initiate and stimulate research. Several of the contributions have later been published, in proceedings and in academic journals (Brändström & Tedebrand, 1988; Sundin & Söderlund, 1977; Tedebrand, 2000).<sup>1</sup> Since 2015 the DDB is part of CEDAR, the Centre for Demographic and Ageing Research, at Umeå University.

## 2.2 SOURCES WELL SUITED TO LONGITUDINAL STUDIES

It is no coincidence that Swedish researchers were among the first to set out to build large, longitudinal databases with individual-level data. What Egil Johansson discovered in the mid-1960s was that the Swedish parish records, with respect to their outline, detail, and coverage, were even better suited for longitudinal research than the sources used by Louis Henry, Peter Laslett, and others. Parish records, including births, marriages, and deaths, as well as family-based continuous registers covering the entire population, had been kept since the late 17th century, and their long time spans offered virtually unparalleled possibilities for longitudinal studies (Nilsdotter-Jeub, 1993). The parish records were not only kept for ecclesiastic purposes; until 1990, they also served as the system of official registration in Sweden. This was a consequence of the historically rooted religious conformity with a national state-church system that until the 1950s included practically all citizens. If someone left the Lutheran State Church they still remained in the church registers, in their capacity as the official system of national registration. This means that the Swedish church registers contain a nearly unsurpassed coverage of the entire population, irrespective of religious affiliation and/or denomination.

A distinctive feature of the Swedish system of national registration is the household-based longitudinal parish registers, or catechetical registers as they also are called, found only in Sweden and Finland. These were dynamic, in the sense that they were continuously updated when new demographic events occurred. The registers also include numerous other forms of personal information data, for example vaccination, reading ability, disabilities, and general conduct. Like in a census, families were kept together on the same page as long as they lived together. When a child was born, a parent died, or a widow remarried, the local minister meticulously recorded this in the register. There is also detailed information about the changing composition of the household group, even for servants changing their positions every now and then. Record

1 Some examples are the conferences/workshops *Time, Space and Man*, Umeå 1977; *Society, Health and Population during the Demographic Transition*, Umeå 1986; *Sex, State and Society*, Umeå 1998; *Genetics, Genealogies and Family Databases*, Umeå 2003; *The Practice of Birth Control and Historical Fertility Change*, Umeå 2008; and *Death Clustering: Towards New Explanations for Infant and Child Mortality in the European Past*, Umeå 2010. The contributions from the last two conferences have been published in *The History of the Family* (2010, 15(2)) and *Biodemography and Social Biology* (2012, 58(2)).

linkage is facilitated by clever links to the event registers, in terms of volume and page, where first-hand information about a particular birth, marriage, or death can be found. The event registers are linked in a similar way to the longitudinal registers, creating a comprehensive system of information whereby individuals can be followed over their entire life spans. With this kind of double bookkeeping, it is also possible to fairly accurately reconstruct missing or lost information. When a longitudinal register volume was completed after five to ten years a new one was established, into which the minister transferred current information from the old volume, of course with valuable links between the old and new registers. For obvious reasons, the longitudinal registers are extremely valuable for the creation of life-course data, making it possible to follow individuals and families over their entire life spans and over generations, without gaps and interruptions, as long as they remain within the parish borders.

The systematic bookkeeping at the local level has made it both practical and natural to carry out the digitisation in the same manner: book-by-book and record-by-record for one parish at a time. The DDB has long had the rare advantage of having a fully employed data entry staff, which has been a great asset in the continuous work, assuring quality and consistency of the data. Their long-term experience is particularly important when handling complicated sources, like very old or damaged books, requiring both skill and proper understanding of the content. At present, the POPUM database offers access to linked data from >100 Swedish parishes, information on approximately 650,000 individuals, and five million records, from the period 1700–1900. The POPLINK database, which is actually a subset of the data in POPUM extended until around 1950 (currently 15 parishes situated in the county of Västerbotten), has linked data on approximately 460,000 individuals and 3.3 million records. New data is continually added to both databases.

### 2.3 STRATEGIES AND DEVELOPMENT

In 1973, the new database project was presented in the leading Swedish historical journal *Historisk Tidskrift* (Johansson & Åkerman, 1973). Although the concept of 'research infrastructure' had hardly been invented, there is no doubt that the new venture had such ambitions. The objective was to, through a joint effort with the Swedish National Archives and the National Labour Market Board, 'produce a resource designated to serve and support research, in Sweden and internationally' (Johansson & Åkerman, 1973, p. 406). The project was to be research-driven, with a clear vision that the database should be built according to sustainable principles, providing a flexibility and versatility that would stand the test of time. Today, almost 50 years later, the majority of these basic principles formulated in 1973 are still relevant in the DDB's data production process. They are as follows:

1. The database shall be *true to the source*. It has to be possible to trace all records back to the original source for verification. The database shall be *complete*; that is, all relevant information in the original source shall be included in the data collection.
2. The data collection shall be *open*, which means that the database shall be built in a way that allows the inclusion of new data, in time as well as in space.
3. The database shall be *coherent* and *consistent*: data entry shall be performed according to similar rules and principles, for maximum comparability and coordination.
4. Data entry, processing, and storage shall be performed in an *efficient* way.
5. All processing of data shall be *research-oriented*, allowing for micro-historic research as well as large-scale cohort studies.

With some modifications, these fundamental guidelines from the early years have also been recognised by others. They are also largely consistent with the core principles proposed by Mandemakers and Dillon (2004) as best practice in the field of building longitudinal historical databases for research.

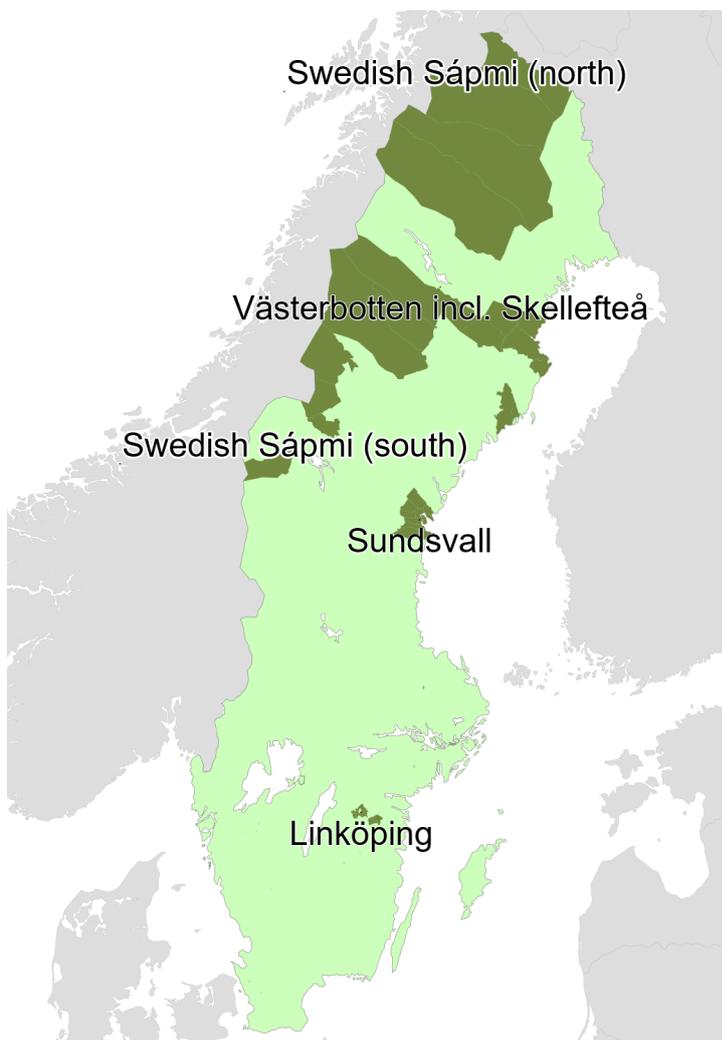
As a pioneer, the DDB has not always had the possibility to lean on golden rules and standards. Its development work was long characterised by learning by doing, and over the years the principles and methods for database-building have frequently had to be reviewed, developed, and sometimes reconsidered. A changing scientific field, new information and communication technologies, and new and advanced methods for analysis have also been important drivers in improving the processes as well as the data. In 1973, the database construction began with the digitisation of parish records from Tuna, a small parish in the industrialised Sundsvall area, which at that time was already an object of interest among historians and social scientists (Brändström, 2009). The work continued with parish records from six other parishes in different Swedish regions, selected on scientific merits. Together, these seven parishes formed the beginning of the POPUM database. Data entry was essentially a manual process, with transcription performed on cards. The

manually transcribed information was later entered into a data terminal. Quite soon, it became clear that working with geographically isolated parishes, however scientifically motivated, was not the best method for database-building. Several of the selected parishes were quite small, and due to in- and out-migration it was difficult to follow individuals over time and over the life course. To improve the coverage and the prospects for life-course studies, since then a conscious and successful strategy has been to collect data from contiguous groups of parishes, or regions, instead of focusing on single, geographically isolated parishes. Covering a cluster of neighbouring parishes usually captures a large number of short-distance migrations and administrative changes over time, increasing the number of complete life courses in the data.

There have, of course, been important methodological developments since 1973, although the basic principles have largely remained unchanged. Manual excerpts, punch cards, and magnet tapes are long gone, and today the data production process is entirely computerised, with some steps being completely automated. Data entry, still a manual process, is performed using custom-made software and, along with the advances of computer technology, manual linkage procedures have been replaced by advanced semi-automatic and automatic record linkage (Larsson & Engberg, 2015, 2016).

The first large region included in POPUM was the industrialised Sundsvall area, a previously agricultural district that in the 19th century became the heart of the sawmill industry in Europe and a centre for the Swedish labour movement. Several important studies have been conducted on the Sundsvall data, illuminating the Swedish industrialisation process and its consequences for individuals, families, and households. In the early 1980s, the Skellefteå area in the north was selected for a large project in genetic epidemiology, which was the start of the digitisation of the parishes in Västerbotten, which now constitute the core population in the POPLINK database. The third region selected for digitisation was Linköping, a socially diverse area in southern Sweden. The fourth large region in POPUM consists of parishes in Swedish Sápmi, with a Sámi and a settler population. Several of these parishes have also been included in POPLINK.

Map 1 Swedish regions included in databases POPUM and POPLINK

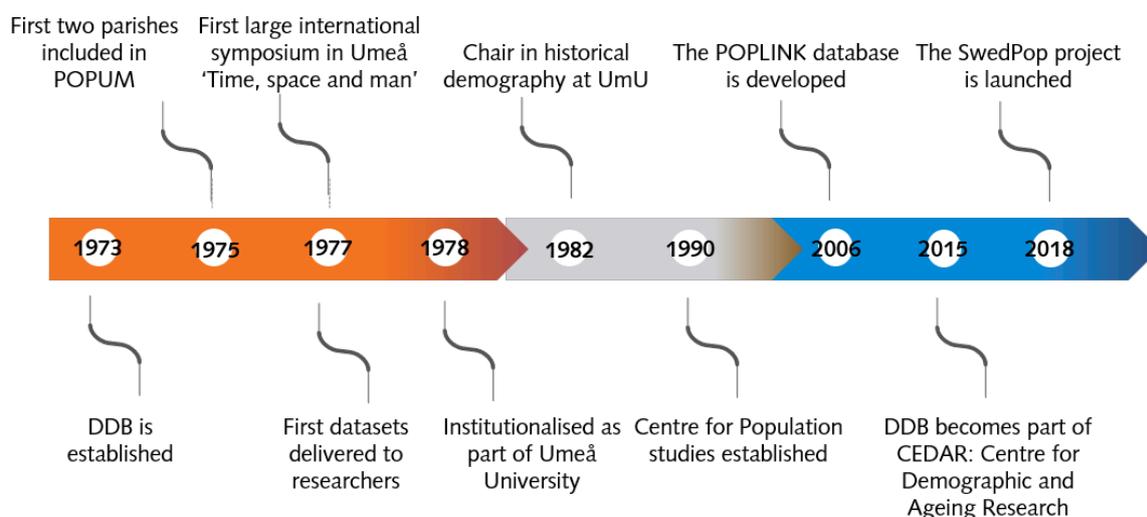


As important as the database-building and management is the dissemination of data. As mentioned above, data can be accessed by all academic researchers, both nationally and internationally, regardless of affiliation. The data in POPUM is open, while the data in POPLINK covering a much later period (until 1950) is surrounded by certain restrictions protecting privacy and confidentiality. Alongside the development of the databases, an organisation for efficient and comprehensive user support and service has been built to help researchers get access to data for their projects. From the very first contact, the individual user receives guidance and assistance from experienced staff, with the aim of delivering a sample that gives the researcher the best possible basis for performing analyses. As the variety of variables and information frequently makes the retrievals complex, for the time being all datasets are prepared at the DDB and delivered to researchers, accompanied by detailed documentation. Although the dissemination of data lies at the core of the DDB's commitment, at times it can also be a challenge for the organisation, particularly when requested data retrievals can be complicated and time-consuming to prepare, and costs are not always fully covered by user fees. Hence, we are continuously working to streamline the processes, without compromising the quality or accessibility. For more information on how to access data from POPUM and POPLINK, see <https://www.umu.se/en/centre-for-demographic-and-ageing-research/order-data/>.

## 2.4 LOOKING TOWARDS THE PRESENT

The most recent development of the DDB databases has been an expansion forward in time, taking full advantage of the long time spans of the rich Swedish sources. The POPLINK database provides access to population data from 15 parishes in the county of Västerbotten from around 1700–1950, with genealogical links covering up to 15 generations. Via a linkage to official registers such as those at Statistics Sweden and the National Board of Health and Welfare, data can be extended into the present time, facilitating large-scale multigenerational studies of present-day cohorts and populations. POPLINK was created from a subset of parishes in the POPUM database, supplemented with new data from the period 1900–1950. In response to the legal framework protecting privacy and confidentiality, we have chosen a safe and flexible model by which additional data, biobank data, research registers, and other contemporary data are not permanently linked to the POPLINK database; these registers remain with each data owner, and are linked to the population data for each project. This is possible thanks to the presence of persistent identifiers (civil registration numbers) in almost all Swedish registers since 1950. Linkage methods have been developed in collaboration with Statistics Sweden, where the linkage is also done. Data retrievals are pseudonymised, and a vetting procedure at the DDB (and sometimes also by an ethical review board) is required. POPLINK and its features, along with methods for linkage, are described in more detail in Westberg, Engberg, and Edvinsson (2016). The project has brought about valuable new collaborations with researchers within both the life sciences and the social sciences, studying intergenerational transfers and risk factors for disease over generations.

Figure 1 *Our history*



A new and important direction in the strategic development of the DDB databases is the increased and intensified collaboration with other Swedish databases. With funding from the Swedish Research Council for the period 2018–2022, the population databases in Umeå, Lund, Stockholm, and Gothenburg, along with the Swedish Censuses at the Swedish National Archives, will be coordinated, harmonised, and made openly available for research in a new research infrastructure, SwedPop. Until now, data from the DDB-databases have only been available as customised datasets extracted by developers in Umeå. With SwedPop this will change. A web platform will be created, where researchers can download harmonised data from all databases, including DDB data, in IDS format as well as rectangular formats. SwedPop will significantly improve access to Swedish population data, and in the long run will also create new prospects for cutting-edge research within the social sciences and historical demography.

## 2.5 METHODOLOGICAL DEVELOPMENTS

In terms of methodology, there have been significant advancements in statistical methods used for analysis. In the first reports and articles from the late 1970s and early 1980s, there was a predominance of descriptive statistics and crude demographic measurements. Since then, along with advances within the computing sciences and the development of new software, the scientific potential of the data has increased through the implementation of new statistical methods. In this context, it can be mentioned that some of the first dissertations in history in Sweden using Event History Analysis were based on data from the DDB (Brändström, 1984; Edvinsson, 1992). Today, this is a standard method for analysing this kind of data. Also worth mentioning is statistician Göran Broström's pioneering methodological work to develop social science applications of Event History Analysis with the software R, using DDB data as real-life examples (Broström, 2012). A recent addition to the methodology is the adoption of sequence analysis, focusing on events and actions in their temporal context, and making it possible to study multiple events and their sequential continuity or change in the same analysis. Complete life courses and individual trajectories are analysed as separate threads, providing information about life patterns, based on similarities in chosen dimensions between sequences, when it comes to the timing and order of states. Sequence analysis has proven highly valuable for studies using life-course data, providing a deeper and more holistic view of changes and transitions throughout the life course (Abbot 1995; Aisenbray & Fasang, 2010; Svensson, Lundholm, de Luna, & Malmberg, 2015; Vikström, Haage, & Häggström Lundevaller, 2017).

The DDB databases POPUM and POPLINK are not only used in quantitative analysis. They can also be analysed using qualitative methods, and several researchers have combined information from the digitised parish registers with complementary data and thus contributed to method development. One example of how to do research on combined data as well as methods through triangularisation has been discussed by Lotta Vikström (2010), exemplified using POPUM data.

Finally, DDB data has proven valuable as a testbed for methodological developments. One example is the development of HISCO (Historical International Standard Classifications of Occupations), where data from the DDB was used in the process of developing an international standard for the classification of occupations within social science research (van Leeuwen, Maas, & Miles, 2002; van Leeuwen, Maas, & Miles, 2004). Data has also been used in the validation of methods for linking Swedish census data and longitudinal population data. By comparing the results of different methods for record linkage of the 1890 and 1900 censuses with the linked digitised parish registers from the Skellefteå regions, the authors were able to show discrepancies between different methods and what might have caused the linkage failures. Without considering the problem of missing surnames of children and family members, the success rate was rather low; but through the inclusion of information on the fathers surname and a created patronym, the success rate improved considerably (Larsson, Berggren & Engberg, 2019; Wisselgren, Edvinsson, Berggren, & Larsson, 2014). For the time being, the DDB is involved in developing a historical classification of causes of death in an international collaborative project (Hiltunen Maltesdotter & Edvinsson, 2020).

## 3 RESULTS AND CONTRIBUTIONS

### 3.1 DATA FOR A BROAD RANGE OF ACADEMIC FIELDS

Looking back at almost 50 years of database-building, we notice considerable development in terms of the size of the database, methodology, and scientific progress. The DDB has become a valuable source for research, for both local scholars at Umeå University as well as external researchers, from other universities in Sweden and abroad. One of the first data retrievals from POPUM in the 1970s was prepared for Roger Schofield, then director of the Cambridge Group for the History of Population and Social Structure. An important step for promoting research on the historical population data was taken in 1982, with the establishment of a Chair in historical demography at Umeå University, with Lars-Göran Tedebrand as the first professor to hold this position. This became very important for developing and strengthening research on the DDB data, as well as for raising a new generation of scholars within the field of historical demography (Tedebrand, 2012). As a consequence of the infrastructural character of the resource, the DDB data has been used in many fields and for many different kinds of research. While there has long been a clear predominance of research within the humanities and social sciences, it has always been used for research in other fields as well, for example medicine.

Over the period from the 1970s to the present, the databases have expanded, in terms of number of records and individuals as well as geographical regions covered. The data has also become more complete, in terms of life courses and genealogies. The addition of new geographical areas opens up for comparative research on different regions, and the inclusion of 20th-century data in POPLINK enables studies of long-term demographic and social development, also covering a central period in Swedish history, the early 20th century, and the establishment of the Swedish welfare state. In all, the expansion in the number of regions, individuals, and complete life biographies has improved its scientific value, and today size and completeness are important prerequisites for the use of the data in research. Size matters!

Before presenting some results based on the DDB data, we would like to direct attention to a number of research perspectives for which historical population databases with microdata such as those at the DDB have resulted in new knowledge made possible by their existence:

- The contextual perspective: The databases allow researchers to study people situated in different historical and spatial contexts. We can reach a deeper understanding of the interplay between individuals and their closest environment.
- The historical perspective: The long time spans allow us to analyse fundamental changes in the lives and living conditions of people from the 17th century onward.
- The family perspective: The family can be used as the unit of analysis and individuals can be studied in their family context.
- The life-course perspective: Individuals and families can be studied from a life-course perspective. The possibility to analyse complete or partial life courses is one of the major advantages of longitudinal historical population databases. Outcomes or choices are understood from previous experiences and conditions.
- The intergenerational perspective: The long time perspective makes it possible to follow families across generations, partly in order to study the historical development of families but also to analyse the influence of conditions and circumstances in previous generations.
- The 'ordinary people' perspective: The advantage of many historical population databases is that they include all social strata in society, i.e. even less affluent groups whose lives we usually lack sources about and who have often been neglected. Differences according to age and gender are typically an integral part of the studies.
- The rare event perspective: An important aspect of the increasing amount of data is the possibility to study infrequent events and conditions, for example rare diseases and disabilities.

In the following, we discuss some central results from studies using data from the DDB, illustrating the perspectives presented above. Egil Johansson's early work demonstrated that the rich — and in many ways unique — Swedish sources, covering the entire population, made it possible to do research on topics that would otherwise not be possible to examine. Access to vital information about fertility, mortality, migration, and marriage in its social context, as well as kinship and family composition over generations for entire populations, became valuable assets for studies in historical demography and

family history. The variety of research is suggested here using a number of examples. While we provide a broad description of the research, presenting many references, this is still merely a selection.

## 3.2 RELIGION, EDUCATION AND CULTURAL CHANGE

The initial studies based on the digitised parish registers in the DDB were within the fields of the history of education and religion. The initiator of the DDB, Egil Johansson, later Professor in Pedagogy, used computerised parish records data from the databases to make groundbreaking contributions to social and cultural history as well as social studies of religion (Graff, Mackinnon, Sandin, & Winchester, 2009). His internationally influential scholarly work has increased our understanding of the social process of alphabetisation, and of the interaction between institutions, families, and the state in the past. The most well-known of his studies are those on the long history of literacy in Sweden since the 17th century (Graff, 2009; Graff et al., 2009; Johansson, 1977). Using information from catechetical registers, where the parish minister noted the ability to read and understand religious texts among parishioners, he documented the programme for making Swedes literate. This began as early as the 17th century, and the marks showed that the majority of the population had become literate already by the late 18th/early 19th century, long before compulsory schooling was introduced in 1842. This had a strong impact on many aspects of Swedish society when it came to participation, adopting innovations, and national identity. However, the ability to write came later. This research field has been continued by several other scholars, most prominently Daniel Lindmark (1995), who has studied how schooling was performed before 1842. This early teaching was based primarily on home schooling, but also on locally organised education (see also Selander, 1986). Despite its origin as church records, the religious homogeneity in Sweden with a national state church makes the data less suited for comparative studies of religion than data on populations in a more diverse religious landscape, such as Canada or the Netherlands. But the information in the parish registers does tell the story of religious change. Bäckström (1999) has studied the change in attendance at catechetical examinations and communion in the late 19th century, a period when the hegemony of the state Lutheran church was challenged. This was expressed in much lower participation in communion and the fact that the yearly catechetical examinations increasingly ceased to be practiced. The grip of the Church lessened, and society became more secularised. He has also studied which groups led the way in this change. Egil Johansson has also studied local society and social differentiation in the early 19th century, based on bench lists and parish registers. These sources tell a great deal about the historical social arrangements and structures. Each family had its designated place, which was based on the formal social division (Johansson, 1983).

An innovative way to study cultural change from the information in parish records is Philologist Linnea Gustafsson's study of naming traditions. She has used data in POPUM to study the introduction of new names in the late 19th century, showing where and in what social groups we find the pioneers of name innovations (Gustafsson, 2002).

## 3.3 DEVELOPING HISTORICAL DEMOGRAPHY

### 3.3.1 MORTALITY STUDIES

In the Umeå group of researchers, mortality research has long had a particular stronghold. As late as the 1980s, there were few mortality studies using individual-level data. Anders Brändström's studies on infant mortality in the northern parish of Nedertorneå were pioneering in several respects, in terms of both methods and perspective. These were the first studies of mortality using computerised DDB data, but also among the first large-scale micro-demographic studies of mortality in Sweden (Brändström, 1984). Brändström's studies made significant contributions to our understanding of the decline in infant mortality in Sweden, and had a large impact on Swedish historical demography. New insights were developed in numerous other studies on infant mortality and the 19th-century mortality decline, which in many cases fundamentally deepened and changed much of the understanding of the Swedish mortality transition (Bengtsson, 1996; Brändström, 1993, 2004; Brändström, Edvinsson, & Rogers, 2000, 2002; Garðarsdóttir, 2002). Brändström's influential work was a study on 19th-century Nedertorneå, a parish close to the Finnish border, where infant mortality was very high. In contrast with what was usually the case, there was no urban penalty. Mortality was higher in the countryside, while children in the small urban part of the parish had better survival. Brändström proved that an almost complete lack of breastfeeding in the traditional rural population explained its high mortality. The tradition of not breastfeeding newborns changed during the 19th century, however, due

to campaigns at different levels in society, from the national medical authorities but also locally through physicians and, not least, midwives. If much of the focus in earlier mortality studies was on economic conditions, Brändström brought in behavioural perspectives. Agency was important, and old traditions could change. Similar aspects have been further discussed by Jan Sundin using DDB data from the Linköping region (Sundin, 1995). The important role of child-care practices in the family and local environments is illustrated by Edvinsson (2004b), who found that in the late 19th-century Sundsvall region, children of mothers migrating from regions with high infant mortality had a significantly less chance of surviving their first year compared to those originating from low mortality regions. Edvinsson interprets this as different traditions of child-care behaviour influencing the newborns' survival chances.

Another major strand of mortality research with DDB data concerns family clustering of deaths and intergenerational aspects of inequality in health and mortality (Edvinsson & Janssens, 2012). The longitudinal registers and high-quality record linkage that allows us to study individuals in their family context over generations makes the Swedish data highly suitable for large-scale studies of different aspects of intergenerational transfers, social and cultural as well as genetic and inherited. Family clustering of deaths was observed already in Brändström's early work. In high-mortality and high-fertility Nedertorneå, some families lost all or almost all their children while in other families all or almost all the children survived (Brändström, 1984). These observations were further developed by Lynch and Greenhouse (1994) and Lynch, Greenhouse, and Brändström (1998); they found a strong intrafamilial correlation of infant mortality in a selection of Swedish parishes, and by analysing mortality according to birth order they reached results suggesting that children of high birth order were not necessarily disadvantaged. A decade later, this theme was again taken up and further developed: Edvinsson et al. (2005) showed that infant mortality in Sweden in the 19th century was highly clustered, with a relatively small number of families accounting for a large proportion of all infant deaths. Two important factors were associated with high-risk families: a biological component, evidenced by an overrepresentation of women who had experienced stillbirths; and a social component, indicated by an increased risk among women who had remarried. The statistical methods of studying clustering have been investigated by Holmberg (2012) and Holmberg and Broström (2012). Another aspect of family difference has been explored by Häggström Lundevaller and Edvinsson (2012), who found that Rh disease contributed to clustering of perinatal mortality in the Skellefteå region.

A related theme is the possibility that mortality patterns are transferred across generations. In a special issue presenting international collaborative studies on the intergenerational transfers of infant mortality, Broström, Edvinsson, and Engberg (2018) found a clear association in infant deaths across generations in the Swedish parishes, a pattern that resembled the results of the other studies (Holmberg, 2012; Vandezande & Edvinsson, 2012). Other examples of intergenerational perspectives in research are presented below, in connection to biological and genetic studies as well as those on social mobility.

Socioeconomic aspects of demography have been a rich field for studies based on DDB data, particularly concerning health and survival. One of the main determinants of survival chances in today's world is social position. Having more economic and social resources and having high social status are strongly associated with better health and lower death risk. While a common assumption is that this has been the case throughout history, studies based on microdata present a more complex picture, showing that the connection between social class and survival is not as straightforward as expected. Already in his early work, Brändström (1984) found that there was no strict social gradient in infant mortality in Nedertorneå. Differences caused by breastfeeding practices were strongly socially determined, but not necessarily in the sense that infant care was better in higher social classes. Edvinsson (1992, 1993) found a much more complex relationship between social class and mortality in 19th-century Sundsvall than what is commonly assumed. The pattern differed between age groups: While only weak associations were found for infant mortality, child mortality (1–14 years) showed a strong disadvantage for working-class children during the industrialisation process. In the adult population, however, the social gradient was not evident; it could even be a disadvantage to belong to a higher social class. The results regarding infant and child mortality have later been further confirmed using a larger dataset for the Sundsvall and Skellefteå regions (Edvinsson, 2004a).

The recent addition of 20th-century data in POPLINK has made it possible to study the long-term development of social inequalities in mortality. In the last decade a couple of studies have been published, focusing on long-term trends in adult and old-age mortality. Surprisingly, the modern social class pattern in mortality has not been confirmed. Instead, the opposite pattern — that high economic and social resources and high social status are associated with higher death risk — has been observed

in different regions with DDB data (Edvinsson & Broström, 2012; Edvinsson & Lindkvist, 2011). Edvinsson and Broström (2017, 2020) show that the present-day pattern of a distinct social gradient in mortality is a surprisingly recent phenomenon among adults and elderly in two regions in northern Sweden. It appears that, in historical Sweden, there was no consistent social gradient in adult and old-age mortality, something that did not appear until the 1970s and 1980s. There is, however, a clear gender dimension in the pattern. Among women there was a consistent gradient with better survival in the higher social classes, while the opposite was the case among men. Other Swedish studies analysing different data find similar patterns (Bengtsson & Dribe, 2011; Bengtsson, Dribe, & Helgertz, 2020; Dribe & Eriksson, 2018). The fact that males and females express different patterns makes the authors hypothesise that gendered and class-specific expectations regarding behaviour played an important role. The gender perspective is taken up by Willner (1999), who studied sex-specific mortality in Linköping. The sex differences were extremely large, pointing at a disadvantage for males in the 19th century, which Willner suggests has to do with behaviours, for example alcohol consumption. The differences decreased rapidly during the first half of the 20th century.

The role of public health and sanitation in the 19th-century urban context has been investigated by Nilsson (1994) using DDB data from Linköping. The last part of the 19th century was a period of increased initiatives and responsibilities for the local governments in cities and towns to improve the poor health conditions in urban environments, and urban mortality decreased considerably in the following decades. Nilsson, studying the role of the local government and the possible impact its initiatives had on health, found that the sanitary improvements had a substantial impact on the improved survival. Edvinsson (1992) found similar results, even if his results showed no clear associations between improved sanitation and better survival on the area level (see also Edvinsson & Nilsson, 1999).

An important aspect of the Swedish mortality transition has been the regional differences, in both levels and trends (Brändström et al., 2000, 2002). One part of this is the disappearance of the urban penalty, partly due to sanitary improvements (see above). Regional differences in economy, schooling, traditions, and social structure became less important, even though they are still strong in Sweden today (and are perhaps even increasing again). That the mortality patterns did not necessarily stop at national borders has been shown by Edvinsson, Garðarsdóttir, and Thorvaldsen (2009) in their study of regional patterns of Nordic infant mortality.

### 3.3.2 DISEASE AND HEALTH RISKS

Quite a good deal of studies have been conducted on specific causes of death, particularly infectious diseases. These studies have provided us with a good overview of the components of the mortality pattern during the mortality transition. They have investigated conditions facilitating the spread of infections, and highlighted the actions and measures to prevent them from spreading as well as the role diseases have had throughout history. Puranen's (1984) thesis is an impressive work, dealing with a wide variety of aspects on tuberculosis in Swedish history. Besides analysing medical knowledge, treatments, and attitudes (their cultural history), she disentangled the epidemiological development of tuberculosis from 1750 to 1950. The disease was already endemic in all parts of Sweden in the 1750s. The regional distribution indicates that there was a close connection between a high incidence of tuberculosis and low living standards. Another thorough and ambitious work on a specific disease is Peter Sköld's (1996a) doctoral study of smallpox in Sweden from the 17th to the 20th century, with a focus on its decline. Sköld found that, although the decline had started as early as the 18th century, the introduction of a vaccination in the early 19th century was effective. The implementation was rapid and commonly accepted, partly due to the use of the Swedish church organisation. He also looked at revaccination and its effects on the life courses of survivors, and found that the disease had long-term consequences, for example on marriage chances (Sköld, 1996a, 1996b).

The life courses of people suffering from venereal diseases in the 19th century have been studied by Anna Lundberg (1999). The spread of syphilis and gonorrhoea was regarded as a major problem. Lundberg investigates the measures taken to deal with the diseases and how they affected the women. She also analysed how the lives of these women developed: mortality was higher, for both themselves and their infants. On the other hand, this likely had more to do with their socioeconomic position and living conditions than with the disease itself.

The incidence and mortality of scarlet fever and diphtheria increased in Sweden during the second half of the 19th century. Curtis (2004, 2008) analysed the spread of these diseases in the Sundsvall region during the period of intense industrialisation in the sawmill district, and how the large influx of migrants to the rural industries could even make the epidemic mortality higher than in the urban environment. He also found that a low nutritional status among mothers during pregnancy made their children more vulnerable to scarlet fever. The history of dysentery and its regional distribution in Sweden has been thoroughly studied by Helene Castenbrandt (2012). This was a very common disease in the 18th century, but declined thereafter.

Historically, one of the major health risks to adult women has been maternal mortality. Using longitudinal data on childbirth and maternal mortality in POPUM, the obstetrician Ulf Högberg (1985, 2004) investigated the development from the 18th to the 20th century, testing hypotheses regarding the determinants of the decline. In a seminal contribution within the field, he found a breaking point towards the end of the 19th century, when risks in connection to childbirth rapidly declined. Högberg's analysis showed a clear connection between the reduced mortality and improved practices of antiseptic methods by midwives at childbirth, a change supported by the emerging welfare state. Andersson, Andersson, Bergström, and Högberg (2000) has further studied this topic. The role of midwives, as mentioned, has been studied by Brändström (1984) and later by Curtis (2005).

Ethnicity has been an important aspect when it comes to health and mortality. Historically, Sweden was largely an ethnically homogenous country. However, parts of Sweden have long been inhabited by ethnic minorities, the most prominent and largest being Finns (many close to the Finnish border) and the Sámi population in the northern inlands — Sápmi. Brändström (1990) has analysed infant mortality among the Sámi population in the northern inlands of the country, finding higher infant mortality compared to the non-Sámi population. In recent years, this theme has been further analysed by Lena Karlsson and colleagues (Karlsson 2013, 2016, 2018; Karlsson, Häggström Lundevaller, & Schumann, 2019). The Sámi had a lower life expectancy, due partly to high infant mortality but also to mortality at other ages. There were clear seasonal differences depending on ethnicity and age group. Sámi infants who were born during winter suffered increased risks of neonatal mortality, while those born during the summer period suffered increased risks of dying after six months of age. The study shows how living in the same environment, but with differences when it comes to culture, economy, and social conditions, led to different patterns of mortality.

The studies of health and mortality among the Sámi relate to a topic that is becoming increasingly important: that of climate and health. Barbara Schumann and others have studied the role of climatic factors for morbidity and mortality in Sweden since the 18th century, using DDB data, looking for factors modifying the impact of climate on health (Karlsson, Häggström Lundevaller, & Schumann, 2020; Oudin Åström, Edvinsson, Hondula, Rocklöv, & Schumann, 2016; Oudin Åström, Forsberg, Edvinsson, & Rocklöv, 2013; Rocklöv, Edvinsson, Arnqvist, Sjöstedt de Luna, & Schumann, 2014; Schumann, Häggström Lundevaller, & Karlsson, 2019).

### 3.3.3 MARRIAGE PATTERNS AND FAMILY STRUCTURE

Fertility, family-building, and family forms in history are also issues that have been addressed using DDB data, with the first studies published in the early 1980s (Larsson, 1984; Lockridge, 1981). In recent years, and with the addition of the POPLINK data, the interest in fertility studies has increased. In a comparative study, Reher, Sandström, Sanz-Gimeno, and van Poppel (2017) studied the propensity to have another child depending on the number of surviving children in the family. Their results indicate a strong role of agency in fertility, with couples regulating their reproduction according to their reproductive goals. Investigating the role of women's socioeconomic status and labour market activity in fertility in Sweden for the period 1900–1960, Glenn Sandström and Emil Marklund (2018) identified the appearance of a strong two-child norm. They also found a decline in a negative socioeconomic gradient of fertility, with white-collar women increasing their fertility more than others.

Among recent studies of fertility is Johan Junkka's (2018a; 2018b; 2018c; Junkka & Edvinsson, 2016) work on fertility patterns and the fertility transition in relation to social networks, in his case membership in voluntary associations and/or the presence of such in the local environment. This could be the free church movement, the temperance movement, or labour organisations. Junkka found that membership and the local presence of such organisations certainly promoted lower fertility, albeit with a somewhat different impact at different times and depending on the type of organisation.

Attitudes, norms, and knowledge were spread within these networks. Junkka's results are very much in accordance with what Paul Roteriing (2017) and Roteriing and Bras (2019) has found in a couple of studies on the Skellefteå region. Roteriing has also focused on the role of networks, but whereas Junkka investigated the role of voluntary associations and the characteristics of the local population, Roteriing considered the family context. He found that there was a clear intergenerational transfer of fertility, and also showed that power relations within the family as measured by the age difference between the spouses played an important role.

The ethnic dimension of fertility and family-building has been studied by Nordin (2009) and Nordin and Sköld (2012). The consequences on the marriage patterns of the meeting between the Sámi population in the north and the non-Sámi population colonising the area during the 18th and 19th centuries were the main object of Nordin's thesis. The Sámi had their own marriage traditions, which partly changed in the cultural meeting with the non-Sámi.

A comparison between patterns of family-building and migration in the two Swedish towns of Sundsvall and Linköping was drawn by Brändström, Sundin, and Tedebrand (2000). Sundsvall was the centre of a strongly developing industrial region in mid-Sweden while Linköping had a different character, that of an administrative centre in the southeast part of the country. However, the marriage pattern in both towns was characterised by strong social and geographical endogamy whereby migrants married migrants and town-born married town-born. These two towns were also compared by Nilsson and Tedebrand (2005), this time with a focus on family-building and the adoption of a new fertility regime, representing new family strategies. Through the use of mostly pre-industrial forms of fertility control as well as spacing and stopping, smaller families were achieved.

During the 20th century, divorce rates increased rapidly in the Western world. Glenn Sandström (2012) has thoroughly studied this in the Swedish context. Using microdata for Västerbotten county in a study of social position (SES) and divorce for the period 1880–1960, Sandström and Stanfors (2020) found a higher divorce rate in higher SES until 1930, which thereafter changed to a negative association when barriers to divorce diminished with industrialisation and modernisation.

Other aspects of family-building and family structure have been investigated by Ann-Kristin Högman (1999) concerning the living conditions of the elderly, and Leonardo Fusè (2008) concerning the proximity of parents to children and whom they transferred their properties to. Maria Bergman (2010) has studied family conditions in sawmill communities, and Inez Egerbladh (1989) peasant households.

### 3.3.4 MIGRATION

Migration to the expanding sawmill area around Sundsvall in the late 19th century has typically been described from primarily a male perspective, with workers searching for jobs in the new industries, for example in a study by Ostergren (1990). But this was just one side of the story. The in-migration to Sundsvall also had a large female component (Vikström, 2001, 2003) as the economic boom created new opportunities for women as well. Analysing the background and life courses of those moving to Sundsvall, Vikström showed that migrants were a highly mobile group but that the migration patterns were strongly gendered. See also Wall (2001) and Nygren (2009).

Migration and mortality have also been a recurrent theme in research using DDB data. Vikström, Marklund, and Sandström (2016) studied the demographic outcomes of colonisation in regions with a Sámi population, and found the Sámi to be a highly vulnerable group. The death rates in this group were higher, and the competition for land made migration rates high. They did find a 'healthy settler effect', however, and the consequences of colonisation were not necessarily ethnically determined. Colin Pooley (2013), comparing the role of locality in Britain and Sweden, found strong similarities in the patterns, as well as the fact that people were closely tied to localities, indicating the impact of family, friends, and community.

## 3.4 SOCIAL HISTORY AND THE STUDY OF MARGINALISED GROUPS IN SOCIETY

One of the main themes of the new social history that became an expanding field in the 1960s and 1970s was giving a voice to ordinary people, and to the poor and marginalised. With their completeness, detail, and richness of information, the Swedish sources have proven to be very well suited to studies of social and economic conditions, and of the lives and opportunities of people in the past. Disabilities,

criminality, and other personal traits were reported in the old parish registers, but this information has long been underused, something that has changed in recent times. In the above-mentioned thesis by Anna Lundberg (1999), she followed the lives of women with sexually transmitted diseases. The lives of unmarried mothers have been investigated by Anders Brändström (1996), who showed that they were certainly vulnerable but were not necessarily outcasts from society. Children of unmarried mothers, however, had much lower chances of surviving their first years (Brändström et al., 2002). By linking individual-level data to other sources, Elisabeth Engberg (2005, 2006) was able to study the determinants of rural poverty and vulnerability in a 19th-century population. The concept of poverty proved to be as complex then as it is today, and was foremost related to different kinds of crises, weak supportive networks, and marginalisation in society. There was also a great extent of lifecycle-associated poverty. Similar methods have been used by Lotta Vikström (2008, 2011) to study the lives of juvenile delinquents in the past. While being exposed to the legal system in youth did not necessarily lead to a criminal life or life in destitution, there were great differences in effects depending on gender. While male delinquents managed fairly well, young females were stigmatised and their life chances were much worse. Expectations depending on gender put women in a different situation compared to men.

Marginalised groups in society were also the focus of Vikström's and Helena Haage's studies on disability. Working with life-course data in POPUM, Haage showed that although marriage chances were smaller and risks of death higher for disabled persons in the 19th century, the disabled were nevertheless a heterogeneous group of individuals with different obstacles and opportunities in life (Haage, 2017; Vikström, Haage, & Häggström Lundevaller, 2017). See also the work on this theme by Olsson (1999) and De Veirman, Haage, and Vikström (2016).

In recent years, Lotta Vikström has continued to investigate the lives of people with disabilities. This is done in an interdisciplinary research group, funded by the ERC (Junkka, Sandström, & Vikström, 2020; Vikström, Haage, & Häggström Lundevaller, 2020; Vikström, Häggström Lundevaller, & Haage, 2017). In this project, disability is studied from many different aspects, in both historical time and contemporary society, and with both quantitative and qualitative sources. POPUM and POPLINK are central in the historical parts of the project. These studies confirm that disability leads to difficult lives for people, but they also document the complexity in this matter. The life courses illustrate both obstacles and possibilities when it comes to jobs, marriages, and a long life. Many could still find decent ways to make a living and build families. This differed substantially between types of disability (sensory/physical/mental), those with mental disorders being the most vulnerable.

### 3.5 POPULATION DATA IN BIOLOGY AND THE LIFE SCIENCES

Over the years, the multigenerational features of the data have also made it useful for research within the life sciences, particularly studies in which access to high-quality genealogies and reliable information on kinship are important. Data from the Skellefteå region was originally included in POPUM for a large project in genetic epidemiology, tracing the origins of Best's macular dystrophy, a rare ophthalmologic disease prevalent in the region (Forsman et al., 1992; Nordström, 1980; Nordström & Thorburn, 1980). Similar data has also been used to study consanguinity and for mapping other diseases (Bittles & Egerbladh, 2005a, 2005b).

An early example of how population data can be useful in studies with a biological approach can be found in a number of studies in the early 1990s by the evolutionary biologist Bobbi Low from the University of Michigan, using DDB data to investigate demographic patterns with a behavioural-ecological approach (Borgerhoff Mulder et al., 2009; Low, 1993; Low, 1994; Low, 2015; Low & Clarke, 1991; Low & Clarke, 1993). A major argument from these studies is that there are certain predictable ecological rules underlying patterns of fertility, mortality, and migration, albeit constrained by a variety of cultural complexities and interactions that cannot be fully measured in aggregate data (Low & Clarke, 1991; Low, Clarke, & Lockridge, 1991; see also Smith, Hanson, & Mineau, 2016). In the last ten years, access to longitudinal biobank data for cohorts and entire populations, along with effective methods for advanced statistical analysis and sequencing, has significantly increased the interest in multigenerational population data among researchers within the fields of medical epidemiology, public health, and genetics. Linking multigenerational population data to biobank samples and registry data makes it possible to follow present-day cohorts back in time and over generations. This offers new perspectives on issues of contemporary relevance, such as risk factors for cardiovascular disease, the interaction of heredity and lifestyle, and various aspects of aging and dementia. For a number of years now there has been an ongoing collaboration between the DDB and researchers in bio-informatics

and genetic epidemiology, developing and evaluating methods for linking and analysing biobank, registry, and population data for studies of risk factors of disease and biological effects in populations. So far this has resulted in two major publications, and more are to come (Kurbasic et al., 2014; Poveda et al., 2016). In a study from 2014, Kurbasic et al. used longitudinal population data in POPLINK, biomedical data, survey data, and prospective biobank data testing methods of imputing genome sequences to study gene-lifestyle interactions in complex diseases. Access to high-quality population data covering many generations has proven to be a valuable asset for this kind of study. Similar data and techniques were used by Poveda et al. (2016), showing that age, sex, and alcohol were likely to be major modifiers of genetic effects for cardiometabolic traits. Another study that can be mentioned in this context is Arslan et al. (2017) showing, based on multigenerational data for four centuries, that fathers of high age had less evolutionary success than others.

### 3.6 SOCIAL MOBILITY AND OCCUPATIONS

The rich information on occupations in the Swedish data and the long time spans make it possible to study historical social structures as well as inter- and intragenerational mobility and marital homogamy. Throughout their lives, occupations and social positions are recorded for adult men. The changing social structure in the expanding Sundsvall district has been shown in several studies, for example Brändström, Sundin, and Tedebrand (2000) and Nilsson and Tedebrand (2005).

Ineke Maas and Marco van Leeuwen have long worked with historical social mobility and social homogamy, and in different geographical settings. In a study of the 19th-century Sundsvall region, Maas and van Leeuwen (2002) investigated total and relative social mobility before and during industrialisation. Intergenerational mobility increased during industrialisation, but stagnated towards the end of the century. Barriers between sectors varied in time and between social groups. Their study indicates a decline in the importance of the transfer of resources, but an increasing importance of education. The Swedish data has also been used by Maas and van Leeuwen in international comparisons. With rich data from many countries' DDB data, they have analysed the long-term development of intergenerational social mobility, and successfully tested the hypothesis that 19th-century industrialisation was the origin of the later increase in mobility. The pre-industrial period was characterised by stable or decreasing mobility, while mobility increased during industrialisation (Maas & van Leeuwen, 2004, 2016).

Van Leeuwen and Maas (2002), in a study on long-term changes in how industrialisation affected homogamy, asked whether a sexual revolution connected to this process led to weakened homogamy through changed preferences among young people and less parental control. Their hypothesis that less control from parents made homogamy weaker was substantiated, while that of an association between industrialisation and weakened homogamy was only partly substantiated. Social and geographical endogamy has also been studied in the article by Brändström, Sundin, and Tedebrand (2000) mentioned earlier. Recently, Kolk and Hällsten (2017) addressed the intergenerational transfers of these issues: They analysed how reproductive success and social mobility in terms of education are transferred over generations, using socioeconomic and demographic information in DDB data from the Skellefteå region. When it comes to both fertility and education, the data suggests effects on lineages into modern times.

A serious problem with the Swedish parish registers as regards analysing social structure and labour in history is the large under-registration of female labour force participation. The family was defined primarily based on the occupation of the head of household, i.e. the husband and father. This easily leads us to underestimate the contribution of women. Lotta Vikström (2010) has compared occupational information from parish registers with that from other sources, and has thus been able to present a much more diverse and nuanced picture of female work. In the local newspapers, women involved in small businesses or other occupations offered their services. Sources such as hospital records also provided better information on female working activities.

## 4 CONCLUSIONS AND PROSPECTS

This has been a quite extensive overview of the history of the DDB and of the research based on DDB data. Still, a great deal of research has not been mentioned, for example studies in economics such as

Sofia Lundberg's work involving past and present public procurement, using data on children boarded out by auction in the 19th century (Lundberg, 2001). This is not to neglect these fields; rather, the selection has focused on the unique character of the microdata in the DDB databases, particularly on the life-course and intergenerational perspectives they offer. We have also aimed to provide examples from different parts of the history of the DDB and from different disciplines, illustrating the wide potential of the data, which can be used for numerous different purposes in a broad range of scientific fields. If we were to characterise the scientific contributions of the DDB data, we could argue that the analyses of microdata sometimes confirmed what was already regarded as common knowledge (but often without really having had the chance to analyse it thoroughly), while in other cases the results were unexpected. The lack of a modern social gradient in mortality during previous centuries is one such example, something that calls for the further development of models for mortality determinants. Also, studies with detailed life-course data revealed a much more complex picture of different phenomena than what has previously been assumed in hegemonic ideas. Furthermore, many of the studies have highlighted the importance of cultural, behavioural, and institutional aspects of demography and social conditions. This reminds us of the need to continuously develop our theoretical models while also taking such aspects into account.

When it comes to future research, it is often difficult to foresee how research infrastructures such as large historical population databases will be used. New opportunities will create new ideas. But we are confident that the most recent increased collaborations, both within Sweden and internationally, will have a fundamental impact on both future research topics and methodological developments. The integration of the Swedish databases within the SwedPop project, makes it possible to analyse central population issues on a national level and perform comparative studies of regional demographic patterns within Sweden. International initiatives such as the HISCO classification system for historical occupations, the development of a new classification of historical death causes within the SHIP project and the IDS standard for structuring and retrieving longitudinal population data substantially facilitate international comparisons. These new opportunities can shed light on a multitude of topics with high relevance for our understanding of the historical development and for the contemporary society, such as the increasing inequality in many countries with serious consequences for health and longevity; the catastrophic consequences of pandemics like the present outbreak of Covid-19, and the increasingly harmful effects of the climate crisis. All these topics can be better analysed and understood with access to rich historical data, telling us about past pandemics, as well as other crises such as famines, natural catastrophes and wars. There is time to consider what can be learnt from these crises, both regarding their causes and their outcomes and management in different historical situations. Finally, we also foresee increased interest from life sciences in long-term and transgenerational data, for example regarding the genetic aspects of diseases and possible intergenerational epigenetic effects.

## ACKNOWLEDGEMENTS

We want to express our gratitude to Umeå University for long term financial support, and to all data entry assistants, system developers, technical and administrative staff, and researchers that over the years have contributed to the development of the Demographic Data Base.

## REFERENCES

- Abbott, A. (1995). Sequence analysis: New methods for old ideas. *Annual Review of Sociology*, 21, 93–113. doi: [10.1146/annurev.so.21.080195.000521](https://doi.org/10.1146/annurev.so.21.080195.000521)
- Aisenbray, A., & Fasang, A. E. (2010). New life for old ideas: The "second wave" of sequence analysis bringing the "course" back into the life course. *Sociological Methods & Research*, 38(3), 420–462. doi: [10.1177/0049124109357532](https://doi.org/10.1177/0049124109357532)
- Alter, G., & Mandemakers, K. (2014). The Intermediate Data Structure (IDS) for longitudinal microdata, version 4. *Historical Life Course Studies*, 1(1), 1–26. Retrieved from <http://hdl.handle.net/10622/23526343-2014-0001?locatt=view:master>

- Andersson, T., Bergström, S., & Högberg, U. (2000). Swedish maternal mortality in the 19th century by different definitions: Previous stillbirths but not multiparity risk factor for maternal mortality. *Acta Obstetrica and Gynecologica Scandinavica*, 79(8), 679–686. doi: [10.1034/j.1600-0412.2000.079008679.x](https://doi.org/10.1034/j.1600-0412.2000.079008679.x)
- Arslan, R. C., Willführ, K. P., Frans, E. M., Verweij, K. J. H., Bürkner, P.-C., Myrskylä, M., Volland, E., Almqvist, C., Zietsch, B. P., & Penke, L. (2017). Older fathers' children have lower evolutionary fitness across four centuries and in four populations. *Proceedings of the Royal Society B. Biological Sciences*, 284(1862). doi: [10.1098/rspb.2017.1562](https://doi.org/10.1098/rspb.2017.1562)
- Bengtsson, M. (1996). *The child at risk: Infant and child mortality in three generations in 19th century Linköping*. Linköping: Linköping University.
- Bengtsson, T., & Dribe, M. (2011). The late emergence of socioeconomic mortality differentials: A micro-level study of adult mortality in southern Sweden 1815–1968. *Explorations in Economic History*, 48(3), 389–400. doi: [10.1016/j.eeh.2011.05.005](https://doi.org/10.1016/j.eeh.2011.05.005)
- Bengtsson, T., Dribe, M., & Helgertz, J. (2020). When did the health gradient emerge? Social class and adult mortality in southern Sweden 1813–2015. *Demography*, 57, 953–977. doi: [10.1007/s13524-020-00877-5](https://doi.org/10.1007/s13524-020-00877-5)
- Bergman, M. (2010). *Constructing communities. The establishment and demographic development of sawmill communities in the Sundsvall district, 1850–1890*. (Dissertation Umeå University). Umeå: Institutionen för idé och samhällsstudier. Retrieved from <https://www.diva-portal.org/smash/get/diva2:345256/FULLTEXT01.pdf>
- Bittles, A. H., & Egerbladh, I. (2005a). Historical patterns of consanguineous marriage in northern Sweden and their influence on the present-day gene pool. Unpublished paper presented at the second international workshop *Genetics, Medicine and History in Brno, 11th–12th May 2005*.
- Bittles, A. H., & Egerbladh, I. (2005b). The influence of past endogamy and consanguinity on genetic disorders in northern Sweden. *Annals of Human Genetics*, 69(5), 549–558. doi: [10.1046/j.1529-8817.2005.00179.x](https://doi.org/10.1046/j.1529-8817.2005.00179.x)
- Borgerhoff Mulder, M., Bowles, S., Hertz, T., Bell, A., Beise, J., Clark, G., Fazzio, I., Gurven, M., Hill, K., Hooper, P. L., Irons, W., Kaplan, H., Leonetti, D., Low, B., Marlowe, F., McElreath, R., Naidu, S., Nolin, D., Piraino, P., Quinlan, R., Schniter, E., Sear, R., Shenk, M., Smith, E. A., von Rueden, C., & Weissner, P. (2009). Intergenerational wealth transmission and the dynamics of inequality in small-scale societies. *Science*, 326(5953), 682–688. doi: [10.1126/science.1178336](https://doi.org/10.1126/science.1178336)
- Broström, G. (2012). *Event history analysis with R*. Boca Raton: Chapman and Hall/CRC Press.
- Broström, G., Edvinsson, S., & Engberg, E. (2018). Intergenerational transfers of infant mortality in 19th-century northern Sweden. *Historical Life Course Studies*, 7(2), 106–122. Retrieved from <http://hdl.handle.net/10622/23526343-2018-0005?locatt=view:master>
- Brändström, A. (1984). "De kärlekslösa mödrarna": Nedgången i spädbarnsdödlighet i Sverige under 1800-talet, med särskild hänsyn till Nedertorneå. (Acta Universitatis Umensis, Umeå studies in the humanities, 62). Umeå: Umeå University.
- Brändström, A. (1990). Från förebild till motbild. Spädbarnsvård och spädbarnsdödlighet i Jokkmokk. In S. Åkerman & K. Lundholm (Eds.), *Älvdal i norr. Människor och resurser i Luledalen 1300–1800* (pp. 307–351). Stockholm: Almqvist & Wiksell International.
- Brändström, A. (1993). Infant mortality in Sweden. In P. P. Viazzo & C. A. Corsini (Eds.), *The decline of infant mortality in Europe, 1800–1950: Four national case studies*. (Historical Perspectives no. 2). Florence: UNICEF.
- Brändström, A. (1996). Life histories of single parents and illegitimate children in nineteenth-century Sweden. *The History of the Family*, 1(2), 205–226. doi: [10.1016/S1081-602X\(96\)90007-X](https://doi.org/10.1016/S1081-602X(96)90007-X)
- Brändström, A. (2004). På återbesök i Nedertorneå. Spädbarnsdödlighet, utomäktenskaplighet och sociala nätverk. In A. Brändström, S. Edvinsson, T. Ericsson, & P. Sköld (Eds.), *Befolkningshistoriska perspektiv: Festskrift till Lars-Göran Tedebrand* (pp. 35–58). Umeå: Umeå University.
- Brändström, A. (2009). Demografiska databasen och historisk demografi i Umeå — "en allvarlig felinvestering"? In R. Jacobsson (Ed.), *Thule: Kungliga Skytteanska Samfundets årsbok 2009* (pp. 211–222). Umeå: Kungliga Skytteanska Samfundet.
- Brändström, A., Edvinsson, S., & Rogers, J. (2000). Infant mortality in Sweden: Creating regions from nineteenth-century parish data. *Historical Methods*, 33(2), 105–114. doi: [10.1080/01615440009598953](https://doi.org/10.1080/01615440009598953)
- Brändström, A., Edvinsson, S., & Rogers, J. (2002). Illegitimacy, infant feeding practices and infant survival in Sweden 1750–1950: A regional analysis. *Hygiea Internationalis*, 3(1), 13–52. doi: [10.3384/hyg.1403-8668.0231](https://doi.org/10.3384/hyg.1403-8668.0231)

- Brändström, A., Sundin, J., & Tedebrand, L.-G. (2000). Two cities. Urban migration and settlement in nineteenth-century Sweden. *History of the Family*, 5(4), 415–429. doi: [10.1016/S1081-602X\(00\)00053-1](https://doi.org/10.1016/S1081-602X(00)00053-1)
- Brändström, A., & Tedebrand, L.-G. (1988). *Society, health and population during the demographic transition. Proceedings of the international conference*. Umeå: Umeå University.
- Bäckström, A. (1999). *När tros- och värderingsbilder förändras. En analys av nattvards- och hushållssedens utveckling i Sundsvallsregionen 1805–1890*. Stockholm: Verbum Förlag.
- Castenbrandt, H. (2012). *Rödsot i Sverige 1750–1900: En sjukdoms demografiska och medicinska historia* (Doctoral dissertation, University of Gothenburg). Retrieved from <http://hdl.handle.net/2077/30195>
- Curtis, S. M. (2004). Nutrition and scarlet fever mortality during the epidemics 1860–90: The Sundsvall region. *Social History of Medicine*, 17(2) 199–221. doi: [10.1093/shm/17.2.199](https://doi.org/10.1093/shm/17.2.199)
- Curtis, S. M. (2005). Midwives and their role in the reduction of direct obstetric deaths during the late nineteenth century: The Sundsvall region of Sweden (1860–1890). *Medical History*, 49(3), 321–350. doi: [10.1017/s0025727300008905](https://doi.org/10.1017/s0025727300008905)
- Curtis, S. M. (2008). In-migration and diphtheria mortality among children in the Sundsvall region during the epidemics of the 1880s. *Journal of the History of Medicine and Allied Sciences*, 63(1) 23–64. doi: [10.1093/jhmas/jrm026](https://doi.org/10.1093/jhmas/jrm026)
- Curtis, S. M. (2011). The arrival and diffusion of academic medicine in rural Sweden: The case of the Sundsvall region in the late nineteenth century. *Hygiea Internationalis*, 10(1) 7–32. doi: [10.3384/hygiea.1403-8668.11101](https://doi.org/10.3384/hygiea.1403-8668.11101)
- De Veirman, S., Haage, H., & Vikström, L. (2016). Deaf and unwanted? Marriage characteristics of deaf people in eighteenth and nineteenth-century Belgium: A comparative and cross-regional approach. *Continuity and Change*, 31(2), 241–273. doi: [10.1017/S0268416016000230](https://doi.org/10.1017/S0268416016000230)
- Demographic Data Base. (2016). *Bibliography 1970–2016*. Umeå: Umeå University.
- Dribe, M., & Eriksson, B. (2018). *Socioeconomic status and adult life expectancy in early 20th-century Sweden: Evidence from full-count micro census data*. (Lund Papers in Economic Demography, No. 1). Lund: Lund University, Department of Economic History. Retrieved from [https://www.ed.lu.se/media/ed/papers/working\\_papers/LPED%202018%201.pdf](https://www.ed.lu.se/media/ed/papers/working_papers/LPED%202018%201.pdf)
- Edvinsson, S. (1992). *Den osunda staden: Sociala skillnader i dödlighet in 1800-talets Sundsvall* (Doctoral dissertation, Umeå University).
- Edvinsson, S. (1993). Urban health and social class. In A. Brändström, & L.-G. Tedebrand (Eds.), *Health and social change. Disease, health and public care in the Sundsvall district 1750–1950* (pp. 55–110). Umeå: Umeå University.
- Edvinsson, S. (2004a). Social differences in infant and child mortality in 19th century Sweden. In M. Breschi & L. Pozzi (Eds.), *The determinants of infant and child mortality in past European populations*. Udine: Forum.
- Edvinsson, S. (2004b). Att ha omsorg om sina barn. Traditionens och miljöns betydelse för spädbarnsdödlighet. In A. Brändström, S. Edvinsson, T. Ericsson, & P. Sköld (Eds.), *Befolkningshistoriska perspektiv: Festskrift till Lars-Göran Tedebrand* (pp. 59–84). Umeå: Umeå University.
- Edvinsson, S., Brändström, A., Rogers, J., & Broström, G. (2005). High-risk families: The unequal distribution of infant mortality in nineteenth-century Sweden. *Population Studies*, 59(3), 321–337. doi: [10.1080/00324720500223344](https://doi.org/10.1080/00324720500223344)
- Edvinsson, S., & Broström, G. (2012). Old Age, health and social inequality. Exploring the social patterns of mortality in 19th century northern Sweden. *Demographic Research*, 26(23), 633–330. doi: [10.4054/DemRes.2012.26.23](https://doi.org/10.4054/DemRes.2012.26.23)
- Edvinsson, S., & Broström, G. (2017). *Life course and long-term perspectives of social inequality in mortality among elderly and adults In Northern Sweden 1801–2013*. Report presented at IUSSP XXVIII, 2017 International Population Conference, Cape Town, South Africa, 29 October–4 November 2017. Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-141173>
- Edvinsson, S., & Broström, G. (2020). Is high social class always beneficial for survival? A study of northern Sweden 1801–2013. (CEDAR Working Papers, No 2). Umeå: Umeå University. Retrieved from <http://umu.diva-portal.org/smash/record.jsf?pid=diva2%3A1174011&dswid=-5834>
- Edvinsson, S., Garðarsdóttir, Ó., & Thorvaldsen, G. (2009). Infant mortality in Nordic countries, 1780–1930. *Continuity and Change*, 23(3), 457–485. doi: [10.1017/S0268416008006917](https://doi.org/10.1017/S0268416008006917)
- Edvinsson, S., & Janssens, A. (2012). Clustering of death in families: Infant and child mortality in historical perspective. *Biodemography and Social Biology*, 58(2), 75–86. doi: [10.1080/19485565.2012.738575](https://doi.org/10.1080/19485565.2012.738575)

- Edvinsson, S., & Lindkvist, M. (2011). Wealth and health in 19th century Sweden. A study of social differences in adult mortality in the Sundsvall Region. *Explorations in Economic History*, 48(3), 376–388. doi: [10.1016/j.eeh.2011.05.007](https://doi.org/10.1016/j.eeh.2011.05.007)
- Edvinsson, S., & Nilsson, H. (1999). Swedish towns during industrialization. *Annales de Démographie Historique*, 2, 63–96. doi: [10.3406/adh.2000.2168](https://doi.org/10.3406/adh.2000.2168)
- Egerbladh, I. (1989). From complex to simple family households: Peasant households in northern coastal Sweden 1700–1900. *Journal of Family History*, 14, 241–264. doi: [10.1177/036319908901400305](https://doi.org/10.1177/036319908901400305)
- Engberg, E. (2005). *I fattiga omständigheter. Fattigvårdens former och understödstagare i Skellefteå socken under 1800-talet* (Doctoral dissertation, Umeå University). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-503>
- Engberg, E. (2006). Life on the edge: Rural poverty and vulnerability in XIXth century northern Sweden. *Annales de Démographie Historique*, 111(1), 31–57. doi: [10.3917/adh.111.0031](https://doi.org/10.3917/adh.111.0031)
- Forsman, K., Graff, C., Nordström, S., Johansson, K., Westermark, E., Lundgren, E., Gustavson, K. H., Wadelius, C., & Holmgren, G. (1992). The gene for Best's macular dystrophy is located at 11q13 in a Swedish family. *Clinical Genetics*, 42(3), 156–159. doi: [10.1111/j.1399-0004.1992.tb03229.x](https://doi.org/10.1111/j.1399-0004.1992.tb03229.x)
- Fusè, L. (2008). *Parents, children and their families: Living arrangements of old people in the XIX century, Sundsvall region, Sweden* (Doctoral dissertation, Umeå University). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-1638>
- Graff, H. J. (2009). Introduction to the history of literacy. In H. J. Graff, A. Mackinnon, B. Sandin & I. Winchester (Eds.), *Understanding literacy in its historical context: Socio-cultural history and the legacy of Egil Johansson* (pp. 14–22). Lund: Nordic Academic Press.
- Graff, H. J., Mackinnon, A., Sandin, B., & Winchester, I. (Eds.). (2009). *Understanding literacy in its historical context: Socio-cultural history and the legacy of Egil Johansson*. Lund: Nordic Academic Press.
- Gustafsson, L. (2002). *Novation i norr: Nya dopnamn och namngivningsmönster i Skelleftebygden 1791-1890* (Doctoral dissertation, Umeå University). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-56801>
- Garðarsdóttir, Ó. (2002). Saving the child: Regional, cultural and social aspects of the infant mortality decline in Iceland, 1770–1920 (Doctoral dissertation, Umeå University). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-56811>
- Haage, H. (2017). *Disability in individual life and past society: Life-course perspectives of people with disabilities in the Sundsvall region of Sweden in the nineteenth century* (Doctoral dissertation, Umeå University). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-130333>
- Haage, H., Vikström, L., & Häggström Lundevaller, E. (2017). Disabled and unmarried? Marital chances among disabled people in nineteenth-century northern Sweden. *Essays in Economic and Business History*, 35(1), 207–238. Retrieved from <https://www.ebhsoc.org/journal/index.php/ebhs/article/view/54>
- Häggström Lundevaller, E., & Edvinsson, S. (2012). The effect of Rh-negative disease on perinatal mortality: some evidence from the Skellefteå region, Sweden, 1860–1900. *Biodemography and Social Biology*, 58(2), 116–132. doi: [10.1080/19485565.2012.720450](https://doi.org/10.1080/19485565.2012.720450)
- Hiltunen Maltesdotter, M., & Edvinsson, S. (2020). *Coding literate cause-of-death information originating from Swedish historical parish registers*. Demographic Data Base, Umeå. In manuscript.
- Holmberg, H. (2012). *Generalized linear models with clustered data* (Doctoral dissertation, Umeå University). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-52902>
- Holmberg, H., & Broström, G. (2012). On statistical methods for clustering: A case study of infant mortality, northern Sweden 1831–1890. *Biodemography and Social Biology*, 58(2), 173–184. doi: [10.1080/19485565.2012.720446](https://doi.org/10.1080/19485565.2012.720446)
- Högberg, U. (1985). *Maternal mortality in Sweden* (Doctoral dissertation, Umeå University). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-1866>
- Högberg, U. (2004). The decline in maternal mortality in Sweden. The role of community midwifery. *American Journal of Public Health*, 94(8), 1312–1320. doi: [10.2105/ajph.94.8.1312](https://doi.org/10.2105/ajph.94.8.1312)
- Högman, A.-K. (1999). *Ageing in a changing society: Elderly men and women in urban Sweden 1830–1930* (Doctoral dissertation, Umeå University). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-54194>
- Johansson, E. (1969/1970). Kvantitativa studier av alfabetiseringen i Sverige. Exempel på källmaterial, metoder och resultat. *Historielärarnas Förenings årsskrift*, 78–93.

- Johansson, E. (1977). *The history of literacy in Sweden, in comparison with some other countries*. (Educational Reports, No. 12). Umeå: Umeå University. Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-62939>
- Johansson, E. (1983). *Kyrkböckerna berättar*. Stockholm: Liber Förlag.
- Johansson, E., & Åkerman, S. (1973). "Faktaunderlag för forskning. Planering av en demografisk databas". *Historisk tidskrift*, 3, 406–414
- Junkka, J. (2018a). *Shared practices. Social networks and fertility decline during the Swedish demographic transition, 1850–1950* (Doctoral dissertation, Umeå University). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-147281>
- Junkka, J. (2018b). Voluntary associations and net fertility during the Swedish demographic transition. *European Journal of Population*, 34, 819–848. doi: [10.1007/s10680-018-9465-5](https://doi.org/10.1007/s10680-018-9465-5)
- Junkka, J. (2018c). Membership and presence of voluntary associations during the Swedish fertility transition, 1880–1949. *Historical Life Course Studies*, 5, 3–36. Retrieved from <http://hdl.handle.net/10622/23526343-2018-0012?locatt=view:master>
- Junkka, J. (2019). Spatial diffusion of fertility decline in Northern Sweden. *Annales de Démographie Historique*, 138(2), 83–117. doi: [10.3917/adh.138.0083](https://doi.org/10.3917/adh.138.0083)
- Junkka, J., & Edvinsson, S. (2016). Gender and fertility within the free churches in the Sundsvall region, Sweden 1860–1921. *History of the Family*, 21(2), 243–266. doi: [10.1080/1081602X.2015.1043929](https://doi.org/10.1080/1081602X.2015.1043929)
- Junkka, J., Sandström, G., & Vikström, L. (2020). The emergence of social gaps in mental health: A longitudinal population study in Sweden, 1900–1959. *PLoS ONE*, 15(4): e0232462 doi: [10.1371/journal.pone.0232462](https://doi.org/10.1371/journal.pone.0232462)
- Karlsson, L. (2013). Indigenous life expectancy in Sweden 1850–1899: Towards a long and healthy life? *Demographic Research*, 28(16), 433–456. doi: [10.4054/DemRes.2013.28.16](https://doi.org/10.4054/DemRes.2013.28.16)
- Karlsson, L. (2016). Advanced ages at death in Sápmi during the eighteenth and nineteenth centuries: With special attention to longevity among the Sami population. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 49(1), 34–49. doi: [10.1080/01615440.2015.1033581](https://doi.org/10.1080/01615440.2015.1033581)
- Karlsson, L. (2018). Indigenous infant mortality by age and season of birth, 1800–1899: Did season of birth affect children's chances of survival? *International Journal of Environmental Research and Public Health*, 15(1). doi: [10.3390/ijerph15010018](https://doi.org/10.3390/ijerph15010018)
- Karlsson, L., Häggström Lundevaller, E., & Schumann, B. (2019). Season of birth, stillbirths, and neonatal mortality in Sweden: The Sami and non-Sami population 1800–1899. *International Journal of Circumpolar Health*, 78(1). doi: [10.1080/22423982.2019.1629784](https://doi.org/10.1080/22423982.2019.1629784)
- Karlsson, L., Häggström Lundevaller, E., & Schumann, B. (2020). Neonatal mortality and temperature in two northern Swedish rural parishes, 1860–1899: The significance of ethnicity and gender. *International Journal of Environmental Research and Public Health*, 17(4). doi: [10.3390/ijerph17041216](https://doi.org/10.3390/ijerph17041216)
- Kolk, M. & Hällsten, M. (2017). Demographic and educational success of lineages in northern Sweden. *Population and Development Review*, 43(3), 491–512. doi: [10.1111/padr.12091](https://doi.org/10.1111/padr.12091)
- Kurbasic, A., Poveda, A., Chen, Y., Ågren, Å., Engberg, E., Hu, F. B., Johansson, I., Barroso, I., Brändström, A., Hallmans, G., Renström, F., & Franks, P. W. (2014). Gene-lifestyle interactions in complex diseases: Design and description of the GLACIER and VIKING studies. *Current nutrition reports*, 3(4), 400–411. doi: [10.1007/s13668-014-0100-8](https://doi.org/10.1007/s13668-014-0100-8)
- Larsson, M. (1984). *Fruktansamhetsmönster, produktionsstruktur och sekularisering: En jämförelse mellan 69 härader vid 1800-talets slut* (Doctoral dissertation, Stockholm University).
- Larsson, M., Berggren, M., & Engberg, E. (2019). Linkage of multiple Swedish Censuses: An evaluation. Unpublished paper presented at the 44th Annual Meeting of the Social Science History Association in Chicago, 2019.
- Larsson, M., & Engberg, E. (2015). *Automatic record linkage over administrative borders: New challenges*. Unpublished paper presented at the 40th Annual meeting of the Social Science History Association in Baltimore, 2015.
- Larsson, M., & Engberg, E. (2016). *How much do link metrics matter?* Unpublished paper presented at the 41st Annual Meeting of the Social Science History Association in Chicago, 2016.
- Laslett, P. (1965). *The world we have lost*. London: Methuen.
- Lindmark, D. (1995). *Uppfostran, undervisning, upplysning: Linjer i svensk folkundervisning före folkskolan* (Doctoral dissertation, Umeå University).
- Lockridge, K. A. (1981). *The fertility transition in Sweden. A preliminary look at smaller geographic units, 1855–1890*. (Report from the Demographic Data Base, No. 3). Umeå: Umeå University.
- Low, B. S. (1993). Resources and reproduction: What hath the demographic transition wrought? *Behavioral and Brain Sciences* 16(2), 300. doi: [10.1017/S0140525X00030120](https://doi.org/10.1017/S0140525X00030120)

- Low, B. S. (1994). Men in the demographic transition. *Human Nature*, 5, 223–253. doi: [10.1007/BF02692153](https://doi.org/10.1007/BF02692153)
- Low, B. S. (2015). *Why sex matters: A Darwinian look at human behavior*. Revised edition. Princeton: Princeton University Press. doi: [10.1002/ajhb.22861](https://doi.org/10.1002/ajhb.22861)
- Low, B. S., & Clarke, A. L. (1991). Family patterns in 19th century Sweden: Impact of occupational status and landownership. *Journal of Family History*, 16(2), 117–138. doi: [10.1177/036319909101600202](https://doi.org/10.1177/036319909101600202)
- Low, B. S., & Clarke, A. L. (1993). Historical perspectives on population and environment: Data from 19th century Sweden. In G. Ness & W. Drake (Eds.), *Population-environment dynamics: Proceedings of the international symposium on population and environment* (pp.195–224). Ann Arbor: University of Michigan Press.
- Low, B. S., Clarke, A. L., & Lockridge, K. A. (1991). *Family patterns in nineteenth-century Sweden: Variations in Time and Space*. (Report from the Demographic Data Base, No. 6). Umeå: Umeå University.
- Lundberg, A. (1999). *Care and coercion: Medical knowledge, social policy and patients with venereal disease in Sweden 1785–1903* (Doctoral dissertation, Umeå University). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-15000>
- Lundberg, S. (2001). *Going once, going twice, sold! The economics of past and public procurement in Sweden* (Doctoral dissertation, Umeå University). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-73622>
- Lynch, K. A., & Greenhouse, J. B. (1994). Risk factors for infant mortality in nineteenth-century Sweden. *Population Studies*, 48(1), 117–133. doi: [10.1080/0032472031000147506](https://doi.org/10.1080/0032472031000147506)
- Lynch, K. A., Greenhouse, J. B., & Brändström, A. (1998). Biometric modeling in the study of infant mortality: Evidence from nineteenth-century Sweden. *Historical Methods*, 31(2), 53–64. doi: [10.1080/01615449809601188](https://doi.org/10.1080/01615449809601188)
- Maas, I., & van Leeuwen, M. H. D. (2002). Industrialization and intergenerational mobility in Sweden. *Acta Sociologica*, 45(3), 179–194. doi: [10.1177/000169930204500301](https://doi.org/10.1177/000169930204500301)
- Maas, I., & van Leeuwen, M. H. D. (2004). Occupational careers of the total male labour force during industrialization: The example of nineteenth-century Sweden. In D. Mitch, J. Brown, & M. H. D. van Leeuwen (Eds.), *Origins of the modern career* (pp. 227–258). Aldershot: Ashgate.
- Maas, I., & van Leeuwen, M. H. D. (2016). Toward open societies? Trends in male intergenerational class mobility in European countries during industrialization. *American Journal of Sociology*, 122(3), 838–885. doi: [10.1086/689815](https://doi.org/10.1086/689815)
- Mandemakers, K., & Dillon, L. (2004). Best practices with large databases on historical populations. *Historical Methods*, 37(1), 34–38. doi: [10.3200/HMTS.37.1.34-38](https://doi.org/10.3200/HMTS.37.1.34-38)
- Nilsdotter-Jeub, U. (1993). *Parish Records: 19th century ecclesiastical registers*. Umeå: Demographic Data Base, Umeå University.
- Nilsson, H. (1994). *Mot bättre hälsa: Dödlighet och hälsoarbete i Linköping 1860–1894* (Doctoral dissertation, Linköping University). Retrieved from <http://liu.diva-portal.org/smash/record.jsf?pid=diva2%3A255928&dswid=5358>
- Nilsson, H. & Tedebrand, L.-G. (2005). *Familjer i växande städer: Strukturer och strategier vid familjebildning i Sverige 1840–1940*. Umeå: Umeå University. Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-13411>
- Nordin, G. (2009). Äktenskap i Sápmi: Giftermålsmönster och etnisk komplexitet i kolonisationens tidevarv, 1722–1895 (Doctoral dissertation, Umeå University). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-22093>
- Nordin, G., & Sköld, P. (2012). True or false? Nineteenth-century Sápmi fertility in qualitative vs. demographic sources. *History of the Family*, 17(2), 157–177. doi: [10.1080/1081602X.2012.687831](https://doi.org/10.1080/1081602X.2012.687831)
- Nordström, S. (1980). Epidemiological studies of hereditary macular degeneration (Best's disease) in Swedish and Swedish-American populations. In A. W. Eriksson, H. Forsius, H. R. Nevanlinna, P. L. Workman & R. K. Norio (Eds.), *Population structure and genetic disorders*. New York: Academic Press.
- Nordström, S., & Thorburn, W. (1980). Dominantly inherited macular degeneration (Best's disease) in a homozygous father with 11 children. *Clinical Genetics*, 18(3), 211–216. doi: [10.1111/j.1399-0004.1980.tb00874.x](https://doi.org/10.1111/j.1399-0004.1980.tb00874.x)
- Nygren, V. (2009). *Mellan två samhällen: Inflyttat arbetsfolk i Linköping under det förindustriella 1800-talet* (Doctoral dissertation, Linköpings University). Retrieved from <http://liu.diva-portal.org/smash/record.jsf?pid=diva2%3A293425&dswid=5358>
- Olsson, I. (1999). *Att leva som lytt: handikappades levnadsvillkor i 1800-talets Linköping*. (Doctoral dissertation, Linköpings University). Retrieved from <http://liu.diva-portal.org/smash/record.jsf?pid=diva2%3A253397&dswid=7863>

- Ostergren, R. C. (1990). *Patterns of seasonal industrial labor recruitment in a nineteenth century Swedish parish: The case of Matfors and Tuna 1846–1873*. Umeå: Umeå University. Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-132355>
- Oudin Åström, D., Edvinsson, S., Hondula, D., Rocklöv, J., & Schumann, B. (2016). On the association between weather variability and total and cause-specific mortality before and during industrialization in Sweden. *Demographic Research*, 35(33), 991–1010. doi: [10.4054/DemRes.2016.35.33](https://doi.org/10.4054/DemRes.2016.35.33)
- Oudin Åström, D., Forsberg, B., Edvinsson, S., & Rocklöv, J. (2013). Acute fatal effects of short-lasting extreme temperatures in Stockholm, Sweden: Evidence across a century of change. *Epidemiology*, 24(6), 820–829. doi: [10.1097/01.ede.0000434530.62353.0b](https://doi.org/10.1097/01.ede.0000434530.62353.0b)
- Pooley, C. G. (2013). The influence of locality on migration: A comparative study of Britain and Sweden in the nineteenth century. *Local Population Studies*, 90(1), 13–27. doi: [10.35488/lps90.2013.13](https://doi.org/10.35488/lps90.2013.13)
- Poveda, A., Chen, Y., Brändström, A., Engberg, E., Hallmans, G., Johansson, I., Renström, F., Kurbasic, A., & Franks, P. W. (2016). The heritable basis of gene–environment interactions in cardiometabolic traits. *Diabetologia*, 60, 442–452. doi: [10.1007/s00125-016-4184-0](https://doi.org/10.1007/s00125-016-4184-0)
- Puranen, B.-I. (1984). *Tuberkulos: En sjukdoms förekomst och dess orsaker: Sverige 1750–1980* (Doctoral dissertation, Umeå University). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-82513>
- Reher, D. S., Sandström, G., Sanz-Gimeno, A., & van Poppel, F. W. A. (2017). Agency in fertility decisions in western Europe during the demographic transition: A comparative perspective. *Demography*, 54, 3–22. doi: [10.1007/s13524-016-0536-0](https://doi.org/10.1007/s13524-016-0536-0)
- Rocklöv, J., Edvinsson, S., Arnqvist, P., Sjöstedt de Luna, S., Schumann, B. (2014). Association of seasonal climate variability and age-specific mortality in northern Sweden before the onset of industrialization. *International Journal of Environmental Research and Public Health*, 11, 6940–6954. doi: [10.3390/ijerph110706940](https://doi.org/10.3390/ijerph110706940)
- Rosenthal, P. A. (2003). The novelty of an old genre: Louis Henry and the founding of historical demography. *Population*, 58(1), 97–130. doi: [10.2307/3246646](https://doi.org/10.2307/3246646)
- Rotering, P. (2017). Intergenerational transmission of reproductive behavior in Sweden, 1850–1889. *Historical Life Course Studies*, 4, 181–202. Retrieved from <http://hdl.handle.net/10622/23526343-2017-0011?locatt=view:master>
- Rotering, P., & Bras, H. (2019). The age difference between spouses and reproduction in 19th century Sweden. *Demographic Research*, 41(37), 1059–1090. doi: [10.4054/DemRes.2019.41.37](https://doi.org/10.4054/DemRes.2019.41.37)
- Sandström, G. (2012). *Ready, willing, able: The divorce transition in Sweden 1915–1974* (Doctoral dissertation, Umeå University). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-60216>
- Sandström, G., & Marklund, E. (2018). A prelude to the dual provider family — The changing role of female labor force participation and occupational field on fertility outcomes during the baby boom in Sweden 1900–60. *History of the Family*, 24(1), 149–173. doi: [10.1080/1081602X.2018.1556721](https://doi.org/10.1080/1081602X.2018.1556721)
- Sandström, G., & Stanfors, M. (2020). *Growing more equal and growing apart? Socioeconomic status and the rise of divorce in Sweden*. (Lund Studies in Economic Demography, No. 4). Retrieved from <https://lup.lub.lu.se/record/c4d0c6a1-1d5d-421d-b963-19946d830ff0>
- Schumann, B., Häggström Lundevaller, E., & Karlsson, L. (2019). Weather extremes and perinatal mortality – Seasonal and ethnic differences in northern Sweden, 1800–1895. *PLoS One*, 14(10): e0223538. doi: [10.1371/journal.pone.0223538](https://doi.org/10.1371/journal.pone.0223538)
- Selander, S.-Å. (1986). *Livslångt lärande i den svenska kyrkoförsamlingen. Fleninge 1820–1890* (Doctoral dissertation, Lund University).
- Sköld, P. (1996a). *The two faces of smallpox: A disease and its prevention in eighteenth and nineteenth-century Sweden* (Doctoral dissertation, Umeå University). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-55842>
- Sköld, P. (1996b). From inoculation to vaccination: Smallpox in Sweden in the eighteenth and nineteenth centuries. *Population Studies*, 50(2), 247–262. doi: [10.1080/0032472031000149336](https://doi.org/10.1080/0032472031000149336)
- Sköld, P. (2010). From inoculation to vaccination: Smallpox in Sweden in the eighteenth and nineteenth centuries. *Population Studies*, 50(2), 247–262. doi: [10.1080/0032472031000149336](https://doi.org/10.1080/0032472031000149336)
- Smith, K. R., Hanson, H. A., & Mineau, G. P. (2016). A six-pack of big ideas for historical demography. In K. Matthijs, S. Hin, J. Kok, & H. Matsuo (Eds.), *The future of historical demography: Upside down and inside out*. Leuven: ACCO Publishers. Available from [https://soc.kuleuven.be/ceso/fapos/files/The-future-of-historical-demogr-KM\\_SH\\_HM\\_JK.pdf](https://soc.kuleuven.be/ceso/fapos/files/The-future-of-historical-demogr-KM_SH_HM_JK.pdf)
- Sundin, J. (1995). Culture, class and infant mortality during the Swedish mortality transition. *Social Science History*, 19(1), 117–145. doi: [10.1017/S0145553200017247](https://doi.org/10.1017/S0145553200017247)

- Sundin, J., Söderlund, E. (1977). *Time, space and man: Essays on microdemography*. Reports from the symposium in Umeå, Sweden. Umeå: Umeå University.
- Svensson, I., Lundholm, E., de Luna, X., & Malmberg, G. (2015). Family life course and the timing of women's retirement – A sequence analysis approach. *Population, Space and Place*, 21(8), 856–871. doi: [10.1002/psp.1950](https://doi.org/10.1002/psp.1950)
- Tedebrand, L.-G. (2000). *Sex, state and society: Comparative perspectives on the history of sexuality*. Stockholm: Almqvist & Wiksell International.
- Tedebrand, L.-G. (2012). *Historiedemografisk forskning vid Umeå universitet 1982–2004: En personlig ämbetsberättelse*. Umeå: Kungliga Skytteanska samfundet.
- Universitets- och högskoleämbetets författningssamling. (1978). *Förordning med bestämmelser för demografiska databasen vid universitetet i Umeå 1978-06-29*, UHÄ-FS 1978 (75).
- Universitets- och högskoleämbetets författningssamling. (1990). *Förordning om demografiska databasen vid universitetet i Umeå 1990-05-03*, UHÄ-FS 1990 (8).
- Vandezande, M., & Edvinsson, S. (2012). Measuring death clustering: Comparing and explaining the spread of infant mortality between families in nineteenth century Belgium and Sweden. In M. Vandezande (Ed.), *Born to die: Death clustering and intergenerational transfer of infant mortality, the Antwerp district, 1846–1905* (pp. 114-143). Leuven: KU.
- van Leeuwen, M. H. D., & Maas, I. (2002). Partner choice and homogamy in the nineteenth century: Was there a sexual revolution in Europe? *Journal of Social History*, 36(1), 101–123. doi: [10.1353/jsh.2002.0111](https://doi.org/10.1353/jsh.2002.0111)
- van Leeuwen, M. H. D., Maas, I., & Miles, A. (2002). *HISCO. Historical International Standard Classification of Occupations*. Leuven: Leuven University Press.
- van Leeuwen, M. H. D., Maas, I., & Miles, A. (2004). Creating a historical international standard for classification of occupations an exercise in multinational interdisciplinary cooperation. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 37(4), 186–197. doi: [10.3200/HMTS.37.4.186-197](https://doi.org/10.3200/HMTS.37.4.186-197)
- Vikström, L. (2001). Women migrants in a Swedish sawmill town: The socio-spatial paths of single women migrants heading for Sundsvall in the 1870s. In L.-G. Tedebrand & P. Sköld (Eds), *Nordic demography in history and present-day society* (pp. 367–384). Umeå: Umeå University. Scandinavian Population Studies. Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-8549>
- Vikström, L. (2003). *Gendered routes and courses: The socio-spatial mobility of migrants in nineteenth-century Sundsvall, Sweden* (Doctoral dissertation, Umeå University). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-14689>
- Vikström, L. (2006). Vulnerability among paupers: Determinants of individuals receiving poor relief in nineteenth-century northern Sweden. *The History of the Family*, 11(4), 223–239. DOI: [10.1016/j.hisfam.2006.12.004](https://doi.org/10.1016/j.hisfam.2006.12.004)
- Vikström, L. (2008). Societal change and individual past in connection with crime: Demographic perspectives on young people arrested in Northern Sweden in the Nineteenth Century. *Continuity and Change*, 23(2), 331–361. doi: [10.1017/S0268416008006814](https://doi.org/10.1017/S0268416008006814)
- Vikström, L. (2010). Identifying dissonant and complementary data on women through the triangulation of historical sources. *International Journal of Social Research Methodology*, 13(3), 211–221. doi: [10.1080/13645579.2010.482257](https://doi.org/10.1080/13645579.2010.482257)
- Vikström, L. (2011). Before and after crime: Life-course analyses of young offenders arrested in 19th century northern Sweden. *Journal of Social History*, 44(3), 861–888. doi: [10.1353/jsh.2011.0001](https://doi.org/10.1353/jsh.2011.0001)
- Vikström, L., Haage, H., & Häggström Lundevaller, E. (2017). Sequence analysis of how disability influenced life trajectories in a past population from the nineteenth-century Sundsvall, Sweden. *Historical Life Course Studies*, 4, 97–119. Retrieved from <http://hdl.handle.net/10622/23526343-2017-0006?locatt=view:master>
- Vikström, L., Haage, H. & Häggström Lundevaller, E. (2020). Marriages among people with disabilities in 19th-century Sweden: Marital age and spouse's characteristics. *The History of the Family*, 25(2), 322–344. doi: [10.1080/1081602X.2020.1719859](https://doi.org/10.1080/1081602X.2020.1719859)
- Vikström, L., Häggström Lundevaller, E., & Haage, H. (2017). First a job, and then a family? Impacts of disabilities on young people's life courses in a nineteenth-century Swedish region. *Disability Studies Quarterly*, 37(4). doi: [10.18061/dsq.v37i4.6095](https://doi.org/10.18061/dsq.v37i4.6095)
- Vikström, L., Marklund, E., & Sandström, G. (2016). Demographic outcomes during colonisation: Migration and mortality among indigenous and non-indigenous populations in nineteenth-century Sweden. *Journal of Migration History*, 2(1), 148–176. doi: [10.1163/23519924-00201006](https://doi.org/10.1163/23519924-00201006)
- Wall, R. (2001). The family circumstances of women migrating permanently or temporarily to Sundsvall in the nineteenth century. *Scandinavian Economic History Review*, 49(3), 46–61. doi:

[10.1080/03585522.2001.10419852](https://doi.org/10.1080/03585522.2001.10419852)

- Westberg, A., Engberg, E., & Edvinsson, S. (2016). A unique source for innovative research: The POPLINK database. *Historical Life Course Studies*, 3, 20–31. Retrieved from <http://hdl.handle.net/10622/23526343-2016-0003?locatt=view:master>
- Willner, S. (1999). *Det svaga könet? Kön och vuxendödlighet i 1800-talets Sverige* (Doctoral dissertation, Linköping University).
- Wisselgren, M. J., Edvinsson, S., Berggren, M., & Larsson, M. (2014). Testing methods of record linkage on Swedish censuses. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 47(3), 138–151. doi: [10.1080/01615440.2014.913967](https://doi.org/10.1080/01615440.2014.913967)

HISTORICAL LIFE COURSE STUDIES  
VOLUME 11 (2021), published 14-09-2021

# The Utah Population Database

## The Legacy of Four Decades of Demographic Research

Ken R. Smith  
University of Utah

Geraldine P. Mineau  
University of Utah

### ABSTRACT

This paper summarizes the unique characteristics of the Utah Population Database (UPDB) and how it has catalyzed demographic, social and medical research since the mid-1970s. The UPDB is one of the world's richest sources of linked population-based information for demographic, genetic, and epidemiological studies at the Individual-level. UPDB has supported hundreds of demographic and biomedical investigations, with heavy emphasis on families, in large part because of its size, representativeness, inclusion of multi-generational pedigrees, and linkages to numerous data sources. The UPDB contains data on over 11 million individuals from the late 18th century to the present. UPDB data represent Utah's population that appear in administrative records and many of these data are updated due to longstanding efforts to add records as they become available including statewide birth and death certificates, hospitalizations, ambulatory surgeries, and driver licenses. The depth of information within UPDB has been used to support a wide range of family, medical and historical demographic studies which are described here arranged into four broad categories: fertility, mortality, life course analyses and some selected special topics. The paper concludes with a discussion of the future areas of innovation within the UPDB and the types of novel studies that they are likely to facilitate.

**Keywords:** Demography, Multigenerational families, Mortality, Fertility, Databases, Life course

DOI article: <https://doi.org/10.51964/hlcs10916>

© 2021, Smith, Mineau

This open-access work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/), which permits use, reproduction & distribution in any medium for non-commercial purposes, provided the original author(s) and source are given credit. See <http://creativecommons.org/licenses/>.

## 1 INTRODUCTION

Understanding the sources of variation in human health and well-being for individuals and families throughout history is a central goal of the social, biological and medical sciences. One strategy for contributing to this understanding is to capture and curate extensive data on large, well-defined populations and all of its members over time. What is most desirable for the research community is to obtain wide-ranging data from 'DNA-to-Demography' or 'Proteins-to-Pedigrees'. Gathering and curating big data for social and health-related research have been successful in several instances, many of which have been fundamental contributors to key medical and social science discoveries. The impact of this data collection strategy will be described in more detail in the following pages but it is noteworthy that organizing historic and contemporary demographic and medical information that incorporate quality controls and centralized management benefit many in the research community. Centralized oversight and careful supervision of data that are made accessible to the research community serve to fuel innovative research with wide ranging impact across many disciplines.

Here, we describe the **Utah Population Database (UPDB)** and specifically its scientific impact, primarily to the social sciences and demography, which offers exceptional and unique data and research opportunities for population scientists, demographers, epidemiologists, geneticists, health services researchers, behavioral scientists, among others, all of whom work on population health and medical research. The distinctive quality of the UPDB is that links individual-level administrative and medical records derived from a range of sources spanning decades for some sources and centuries for others (Casey, Schwartz, Stewart, & Adler, 2016; Hurdle, Smith, & Mineau, 2013). The UPDB is a research resource that has been expanded extensively in its 40 years of existence. At this time, UPDB includes information on over 11 million individuals who have basic demographic information and have been a data source for nearly 400 research projects. The time period covers birth cohorts from the 1700s but are more extensive starting in the mid-1800s through the present. The UPDB is in the same genre as other large data projects in the US and UK such as the ongoing longitudinal Framingham Heart Study and the National Survey of Health and Development which are massive in scope, years of coverage, and depth of information.

### 1.1 CONTENT OF THE UPDB

The visionaries who established UPDB over 40 years ago sought to integrate genetics and the social sciences. Its beginnings, during the years 1973–1974, started when several researchers at the University of Utah realized the research opportunities that could be gained by first obtaining extensive genealogy records and constructing a population-based resource that would link these data to high quality medical records in order to investigate the genetic basis of a number of important diseases. Central to this effort was geneticist Mark Skolnick, who led a consortium of key scientists including, cardiologist Roger Williams and demographer Lee L. Bean.

The original set of genealogy records used when the UPDB was being developed comprised approximately 185,000 documents representing, on each form, three generations: a husband-wife pair, their four parents, and the couple's offspring and their respective spouses. These initial documents were selected to represent approximately 1.9 million individuals. Linking these across generations created thousands of multi-generational pedigrees, providing astonishing insights regarding the population (Song & Campbell, 2017).

These early genealogical records comprise the original backbone of UPDB. The founding research team secured access to the Utah Cancer Registry (UCR, a Surveillance, Epidemiology and End Result [SEER] Registry) and Utah death certificates (from the Utah Department of Health) as the basis for medical outcomes to be linked to the genealogies at the individual level.

While the original development of UPDB was derived from genealogy, cancer and death records, the UPDB now includes substantially more records and from diverse sources (see Table 1). In addition to the original genealogy records, UPDB includes (1) All available Utah vital records (births, deaths, marriages, divorces and fetal deaths), (2) Statewide health facilities data including Inpatient Hospital Discharge, Emergency Department, Ambulatory Surgery records from the Utah Department of Health, (3) Health insurance claims data ('All Payer Claims') (4) Utah cancer registry, (4) Social Security death index records, (6) Utah Voter Registration, (7) Utah driver license records, and (8) The 1880 and 1900–1940 Utah individual-level censuses and other data sets from the Department of Health.

Table 1 Overview of key data sources comprising the Utah Population Database

Record Type	Years Available	Records
Original Family History Records	1700's–1975	1,916,649
Birth Certificates	1915–1921, 1926–present	3,067,098
Death Certificates	1904–present	921,081
Marriage Certificates Utah	1978–2010	692,680
Divorce Records Utah	1978–2010	298,928
Fetal Deaths Utah	1978–present	11,434
Ambulatory Surgery Utah	1996–present	7,255,110
Inpatient Hospital Claims Utah	1996–present	6,124,266
Emergency Department	1996–present (in process)	9,716,843
All Payer Claims Data	2013–present	3,561,630
Utah Cancer Registry	1966–present	394,437
Utah Birth Defect Network	1995–2015	21,235
U.S. Census of Utah	1880,1900–1940	2,300,084
Driver License Utah	Updated annually since 1990	3,997,558
Voter Registration Utah	Updated after national elections since 2008	2,068,916

Two distinctive features of UPDB are noteworthy with respect to linkages to other large medical data sets. First, links have been created connecting UPDB with the data warehouses of the two largest health providers in Utah — University of Utah Health and Intermountain Healthcare (DuVall, Fraser, Rowe, Thomas, & Mineau, 2012). These two health care providers represent inpatient and outpatient electronic medical information for approximately 85% of the state's medical encounters starting from the mid-1990s. The medical data per se are not held within the UPDB but are securely maintained by the data warehouses of the providers. Medical data are joined with the demographic and genealogical data in UPDB after the research project receives the necessary approvals from appropriate Institutional Review Boards and the Utah Resource for Genetic and Epidemiologic Research (RGE), which oversees research access to the UPDB.

A second and related medical data linked to the UPDB are those derived from Medicare claims. The Medicare data are available due to funding from National Institutes of Health (NIH) grants in order to facilitate the study of healthy aging and health expectancy among the Medicare-eligible population. These data are available to researchers using the UPDB but they must not only obtain IRB approval for their use but also approval from the Centers for Medicaid and Medicare Services (CMS).

Linking all the records as shown in Table 1 within the UPDB creates unique research opportunities including:

1. Creation of reproductive histories. Using data from the Utah Department of Health that includes Utah birth certificates from 1915 to the present, genealogical holdings of UPDB have been extended considerably. Mothers and fathers on multiple birth certificates are linked. This allows us to see that specific individuals share common parents and are therefore siblings. The children named on these birth certificates are then linked to the birth certificates of their children. Because birth certificates provide gestational age and birth weight as well as other features such as adverse obstetric events and birth complications, this strategy has provided a valuable source for analysis of obstetric health in families and across generations (Hammad et al., 2020; Theilen et al., 2016; Theilen et al., 2018). Many of the genealogies derived from vital records also link into the legacy genealogies that are part of the UPDB.
2. Creation of residential exposures and histories. Location information is derived from several sources in UPDB including Driver License Division (DLD) data, voter registrations, and vital and birth records. One use of DLD is to determine if an individual is currently under observation, while residence information on death records verify an individual was under observation until their death. Voter Registration records are obtained and linked to UPDB which give location information at a particular

point in time. Additionally, DLD data hold information on height and weight from which Body Mass Index (BMI) for each individual can be derived (Chernenko, Meeks, & Smith, 2019; Smith et al., 2008; Smith et al., 2011; Zick et al., 2009). Finally, residential histories within UPDB have been geo-coded which create the opportunity for linking any geo-referenced data set (e.g., census block and air quality monitors) with individual-level data. In general, for contemporary decades where address information was captured, UPDB generates location data down to the Census block-group level which are statistical divisions of US census tracts and have 600 and 3,000 people. Of course, higher level areas of aggregation are also available. For more historic years, Census Enumeration Districts are provided or place names (city, county).

3. Individual-Level Census Records. The addition of the micro level census records from 1880 and 1900–1940 to UPDB allows for several types of studies. First, it is now possible to observe mobility, both geographic and socioeconomic, and its causes and consequences. Second, given the manner in which census enumerators were assigned to districts to conduct the full count of the population, the data are arranged into neighborhoods, as noted, represented as Enumeration Districts. Accordingly, individuals identified in the census can be characterized by the quality of their 'neighborhoods' and how these spatial attributes may alter later life outcomes. Finally, these census records provide valuable independent information about family composition, co-residence, and genealogical data that may not be possible from other sources of data in the UPDB.
4. High Risk Pedigrees and Gene Discovery. The initial motivation for creating the UPDB stemmed from the concept of joining multigenerational information to medical records in order to facilitate the identification of genes involved in disease risk. With this information, cancer in the early (and present) era of UPDB, and now almost any condition found in electronic medical records, provides scientists with the opportunity to identify families with an excess disease risk (or traits that are positive such as longevity). Family-based studies have an advantage in providing more information from which to identify how genes segregate among related individuals who are or are not affected by the disease. UPDB has provided the family and disease information from which many gene discoveries are based including genes for breast cancer (BRCA1, BRCA2), melanoma (p16), and colon cancer (APC).
5. Link to Existing Cohorts. UPDB also has the capacity to link its data to ongoing projects that have arisen independent of UPDB. For example, the Cache County Memory and Health Study was launched in 1995 to study factors related to dementia and Alzheimer's disease risk. These persons were 65 and older at enrollment and were from a single county in Northern Utah. This linkage between an existing cohort and the UPDB provided an opportunity to open up new life course studies of dementia and Alzheimer's disease (Norton et al., 2010, 2011, 2016).

## 1.2 RESOURCE FOR GENETIC AND EPIDEMIOLOGIC (RGE) RESEARCH

Access to UPDB data is regulated by the Utah Resource for Genetic and Epidemiologic Research (RGE). The RGE was created by an Executive Order of the Governor of Utah in 1982. Relying on enabling statutes in state health code, the RGE was established as a "data resource for the collection, storage, study, and dissemination of medical and related information" to operate "for the purpose of reducing morbidity or mortality, or for the purpose of evaluating and improving the quality of hospital and medical care". Originally administered under the direction and supervision of the Utah Department of Health, the RGE was transferred to the University of Utah by a second Executive Order in 1986. RGE is the legal custodian for the data contained within the UPDB and is responsible for developing and maintaining contractual agreements with organizations that contribute data to the UPDB or that links records to the UPDB.

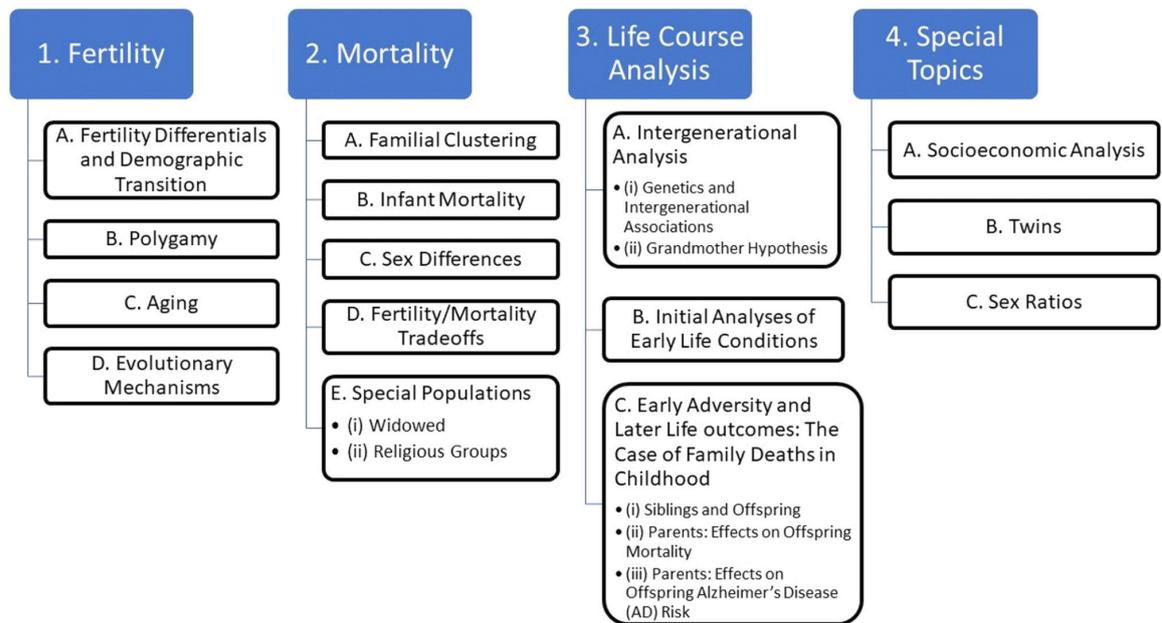
Each project requesting access to data from the UPDB or linked electronic medical records applies to RGE for review. Applications are reviewed by the RGE Committee, which includes University faculty with expertise in several disciplines including demography, genetics, public health and epidemiology, as well as representatives from each of the data contributors. All projects are required to obtain approval by the appropriate Institutional Review Board(s) and Privacy Board(s) before access to data is granted. Ultimately, RGE has the responsibility to protect the sensitive confidential information in UPDB.

## 1.3 IMPACT OF UPDB ON RESEARCH DOMAINS IN DEMOGRAPHY

The birth and maturation of UPDB as a unique data resource was initially developed to promote both demographic and genetics research. The summary of scientific impacts of the UPDB presented here focuses on demographic research and on publications where demographic and genetic elements are combined.

The number of publications in this realm are substantial and we do not attempt to include all publications. Instead, the focus is on publications that exemplify the types of research in four broad demographic domains that represent the value of UPDB to these research endeavors. Understandably, some investigations span more than one of these domains. Accordingly, highlights of a given publication may get attention in more than one section. The organization of the four broad domains and their sub topics are summarized in a roadmap in Figure 1 and this provides the structure of the review that follows.

Figure 1 UPDB Research Domains



## 2 FERTILITY

One of the earliest sets of questions behind the development of the UPDB was born from issues about fertility and how reproduction forms the foundation for the study of genetics and inheritance. Given the settlement history of the Utah territory in the mid-19th century and the migration of adherents to the Church of Jesus Christ of Latter-day Saints, initial research projects focused on the early pattern of natural fertility, and the study of fertility and its change during the demographic transition.

### 2.1 FERTILITY DIFFERENTIALS AND DEMOGRAPHIC TRANSITION

The early years of the UPDB focused on fertility patterns, launched with a two-part analysis of 'Mormon demographic history'. The first examined nuptiality and fertility among once-married couples (Skolnick et al., 1978) and a second on 'the family life cycle and natural fertility' (Mineau, Bean, & Skolnick, 1979). Skolnick and his colleagues motivated their analysis with a recognition that the Church of Jesus Christ of Latter-day Saints had a history of polygamy in the 19th century and endorsed pro-natalist policies all of which led to higher fertility than the nation at large at that time. While these patterns have been known for some time, they argued that the specific characteristics of this fertility behavior in Utah had been undeveloped. In particular, during the early settlement years (in the two decades after 1847, when white settlements in Utah started) marriage ages declined and the proportion of women marrying as teens increased. During the middle of the 19th century, little fertility control was practiced and age at marriage was a major determinant of total and completed fertility. The period 1870–1880 shifted demographic history of Utah when the transcontinental railway opened, exposing Utah residents to greater influence of the secular world. This was also the decade where church members were facing pressure from non-church members in Utah and elsewhere to renounce the practice of polygamy.

The second of these paired papers by Mineau and collaborators (1979) drew on the debate in the 1960's and 1970's about the American family life cycle which was largely based on women born during and after the 1880s and presumably portrayed a 'modern' pattern (i.e., declining and controlled fertility, lower rates

of infant mortality and improved parental life expectancy). They examined a unique American population (Utah) and the opportunity it provided to study the life-cycle under conditions of natural fertility (pre-1870). They extended the study of the family life-cycle to US birth cohorts (1800–1869). In comparisons of the family life cycle at the time, Mineau et al. showed differences in the number of years between the marriage of the last-born child and the age at family dissolution (the 'empty nest' phenomenon). Using micro-data from the UPDB (rather than life table parameters), Mineau et al. concluded that the increasing number of years attributed to the 'empty nest' phenomenon probably has less to do with increased survival rates of parents and more to the interaction between declining fertility levels and decreasing age of mother at birth of the last child.

In a series of papers, Anderton led several studies that provided new insights into fertility limitation and transitions after the early natural fertility settlement years. In 1984, Anderton and colleagues (Anderton, Bean, Willigan, & Mineau, 1984) concluded that during the period of natural fertility, commitment to a pronatalist faith helped to account for differences in fertility levels, but not patterns of fertility decline over time. Factors typically associated with fertility decline in Western Europe such as consistent long-term secularization and urbanization, were found to be more important determinants of cross-sectional fertility levels and longitudinal changes in fertility levels across the birth cohorts studied. This question was extended to consider the role of birth spacing and fertility transitions (Anderton & Bean, 1985). In this analysis they hypothesized that substantial groups of women with long birth intervals could be identified — even during periods when fertility behavior at the aggregate level is consistent with a natural fertility regime. Specifically, they concluded that birth spacing patterns are highly parity-dependent and that the transition is associated with a larger proportion of women shifting to the same spacing schedules associated with smaller families in earlier cohorts. They also showed that changes in birth intervals over time are indirectly associated with age of marriage and that there is evidence of efforts to terminate child-bearing (i.e., shifts in stopping behavior). Overall, they emphasized the importance of distinguishing between spacing and stopping behavior. Later, Hsueh and Anderton (1990) evaluated age, period, and cohort effects on marital fertility during the onset of the Utah fertility transition (1880–1900). They concluded that declining marital fertility in Utah can be explained by both declining fertility levels across historic periods and increasing age-specific limitation across cohorts. They argued that fertility levels were adaptive (through birth spacing across ages) to immediate contexts of childbearing while age-specific fertility truncation increased across cohorts (through diffusion of contraceptive innovations).

These earlier analyses were accompanied by additional foci on central demographic questions regarding fertility. We consider three additional specific domains: the role of polygamy, aging and reproductive senescence, and analyses motivated by evolutionary theories of fitness.

## 2.2 POLYGAMY

In early Utah history, the Church of Jesus Christ of Latter-day Saints supported plural marriages though it was renounced later in the 19th century as a basis for achieving statehood (Ellsworth, 1963; Lyman, 1998). This fact has been the basis for several important historical demographic analyses. Bean and Mineau examined plural marriages in the context of fertility (Bean & Mineau, 1986). They specifically addressed the relationship between polygynous marriages and fertility which the literature tends to show that woman-specific fertility levels are lower in polygynous than in monogamous marriages. Using data for 2,534 polygynists with 7,378 marriages and comparing fertility with once-married monogamous women, they found significant differences in fertility levels by the order of the plural wife. They note that in most polygynous cultures more sister wives marry-in as the husbands increase in status, and therefore husbands will be older at the time the second and later wives are added. With increasing age of the patriarch, widowhood is also more likely so that the risk of pregnancy is affected by wife-order. With UPDB, the average level of fertility of all polygynous wives is lower than that of monogamous wives. This is due to a significantly lower fertility among second and later wives than among monogamous wives, while the fertility of the first wife in polygynous families is higher.

Moorad published a series of papers (Moorad, Promislow, Smith, & Wade, 2011; Moorad, 2013; Moorad & Wade, 2013) using UPDB in which he addresses the role that polygyny plays in sexual selection. Sexual selection is the competition of one sex for reproductive access to the other. Moorad was interested in the hypothesis that sexual selection is stronger in polygamous than in monogamous species. Moorad and his colleagues (2011) found, for example, that over the reproductive lifetimes of Utahns born between 1830–1894, reductions in the rate and degree of polygamy was associated with a 58% reduction in the strength of sexual selection. Polygyny conferred a strong advantage to male fitness (more progeny who survive) as well as a weak disadvantage to female fitness. In contrast, mating with multiple males provided little benefit to females in this population.

## 2.3 AGING

It is axiomatic that the time of marriage and first birth have a profound effect on subsequent fertility. In the US, it has been difficult to show how this pattern plays out in the context of natural fertility. Mineau and Trussell (1982) examined these associations using birth cohorts in UPDB from 1840 to 1879. They showed that older-aged husbands depress marital fertility only at higher marriage durations. They also demonstrated that mother's aging is the most important factor, while father's aging has a moderately negative effect under a natural fertility regime.

The UPDB has been used to compare fertility in relation to other historical databases spanning periods of natural fertility. The most notable example is a large international comparison of reproductive behavior across several populations: The Utah Population Database (UDBP), the *Registre de la population du Québec Ancien* at the Université de Montréal, Canada (RPQA), the LINKS database, hosted by the International Institute of Social History in Amsterdam, the German database based on *Ortssippenbuch* ('book of local kinsmen'), the BALSAC demographic database representing the Saguenay-Lac-St-Jean (SLSJ) region in Québec, and the database collected by the French demographer Louis Henry of the French population between 1670 and 1830 (Eijkemans et al., 2014). The question was whether it was possible to quantify the ages after which women are biologically unable to reproduce? The analysis was motivated by the fact that little is known about the distribution of female age at last birth (ALB) especially now with the widespread availability of modern birth control. The six natural fertility populations comprised nearly 60,000 women. Eijkemans and his colleagues showed that while these populations represent different historical time periods and cultural contexts, the distribution of ages at last birth is quite similar. The curve denoting the end of fertility indicates that <3% of women had their last birth at age 20 years and that about 50% had their age at last birth by age 41, almost 90% by age 45 years and approaching 100% at age 50 years.

## 2.4 EVOLUTIONARY MECHANISMS

The availability of high-quality demographic and life history data spanning decades consistent with natural fertility has attracted the attention of evolutionary biologists and anthropologists. Two analyses are highlighted here that exemplify this line of inquiry.

Jones and Bliege Bird (2013) examined a fundamental hypothesis that argues that fitness, that is reproductive success, should be maximized by an intermediate level of fertility. This prediction has not been widely supported in the human life-history literature and they contend that the difficulty of finding this intermediate reproductive optimum may be a measurement issue. Rather than using lifetime reproductive success as the fitness measure, they proposed a measure that accounts for *variation* in reproductive timing which better reveals preferences about when women are making risky reproductive decision-making. Using UPDB and data from 19th century, they demonstrate that if births are properly timed, a lower-fertility reproductive strategy can have the same fitness as a high-fertility strategy.

Gagnon and collaborators (2009) also considered tradeoffs using frontier populations: UPDB, the *Registre de la population du Québec ancien* (Université de Montréal), and the BALSAC database (Université du Québec à Chicoutimi from the Saguenay-Lac-St-Jean (SLSJ) in Québec. These data provided exceptional opportunities to test the hypothesis of a trade-off between fertility and longevity. Together, these databases allow for comparisons over time and space, and represented one of the largest comparisons of natural fertility cohorts to simultaneously assess reproduction and longevity. They observed a negative influence of parity (more children, worse survival) and a positive influence of age at last birth (ALB; later ALB, better survival) on postreproductive survival in the three populations, as well as a significant interaction between these two variables, patterns that were remarkably similar in the three samples. Strong support was therefore found for a trade-off between parity and longevity and a strong moderating influence of ALB.

## 2.5 FERTILITY SUMMARY

The UPDB has contributed to our understanding of fertility, the forces influencing its variation, and what it tells us about post reproductive consequences. While the volume of studies is considerable, we suggest that UPDB has, in general, provided an opportunity to study in great detail the shifts in fertility for 200 years, at the individual, family, pedigree, and community levels. The historical reach of the data, starting from the early settlement pioneer era, have attracted the attention of anthropologists and biologists, as well as historical demographers, interested in testing evolutionary hypotheses best conducted on humans during years approximating natural fertility conditions. At the other extreme, these historical data connect to present day Utahns such that the potentially enduring effects of past fertility behavior can be assessed.

This feature has led to collaborations with obstetrics and gynecology colleagues who are asking previously underexplored questions that require historical perspectives and data. Some unique components of UPDB have allowed insights into fertility behavior including plural marriage, how demographic change happens when a population moves into a sparsely inhabited geography and how it creates the conditions for a society to grow in the following decades.

## 3 MORTALITY

A powerful advantage of the UPDB is its coverage of key life history and life course traits. The strategy to capture birth, marriage, and death information from a variety of sources including church records, vital registration and medical sources provides an opportunity to study mortality patterns and differentials over time for the Utah population with triangulation across diverse types of data. With death registration beginning in Utah in 1904, UPDB now holds over 100 years of cause of death information all coded to the International Classification of Diseases. Given that mortality information is linked to UPDB's vast genealogical data, many studies have explored familial forces that affect and are affected by the timing and causes of death among family members. Here we summarize selected key contributions of UPDB data applied to questions related to mortality.

### 3.1 FAMILY CLUSTERING

It is well-established that mortality risks are often shared among nuclear and extended family members. The UPDB has been the basis for developing methods which are then applied to increase our understanding of mortality risks beyond individual level measures. One of the earliest and influential methods was introduced by Kerber (1995). The innovation here was to develop a method for estimating excess relative risks of mortality or disease to an individual that could be due to familial factors. Using UPDB, all relatives of a person are identified and measured for the presence and timing of a disease or a cause of death. The count of disease X or cause-of-death Y among the relatives of the person is compared to a count of what would be expected if those relatives had experienced the population risks, given their age and sex. Using exposure years (to generate incidence rates) and adjustments for how closely related a person is to each of their relatives, the method produces a measure similar to a relative risk (which compares observed versus expected rates) called the familial standardized incidence ratio (FSIR) or familial standardized mortality ratio (FSMR). For example, for a study of suicide, a significant FSIR > 1 means that for people who died of suicide, they have relatives who were more likely to die of suicide than would be expected based on the prevailing population suicide rates, an indication of familial clustering.

This technique has been expanded and applied to several mortality questions. Kerber et al. (2001) evaluated the influence of family history of longevity by examining longevity in a cohort of 65+ individuals drawn from UPDB born between 1870–1907. Using the logic of the FSIR applied to longevity, resulting in a measure called Familial Excess Longevity (FEL), they showed that excess longevity aggregates in families, and that the presence of familial aggregation of longevity is a powerful predictor of longevity a given person. O'Brien and colleagues (2007) considered the effects of familial longevity and familial mortality on mortality rates for 10 leading causes of death. FEL and FSMR were estimated for 666,921 individuals over 40 born from 1830 through 1963. They showed that a family history of disease increases the risk of dying from the same cause, whereas a family history of longevity is protective for most age-related diseases including heart disease, stroke, and diabetes, but not cancer (Kerber, O'Brien, Smith, & Mineau, 2008).

The UPDB has been used when applying another measure of familiarity of disease or mortality, the Genealogical Index of Familiarity or GIF (Cannon Albright, 2008). The idea is simply to identify the genetic relationships between all pairs of individuals with the same disease or cause of death and to then estimate the average relatedness of these individuals. This uses the (Malécot) coefficient of kinship to measure the relatedness of each pair of cases between individuals sharing a cause of death, for example. This is repeated for matched controls for the average relatedness one would expect in the general population reflected in UPDB. If the average relatedness of a set of people dying from a given cause is significantly higher than the mean relatedness from a set of matched controls, there is support for excess familiarity for this cause of death though it may represent genetic or environmental forces. GIF have been used to study familial clustering of disease and mortality using UPDB, for example, Alzheimer's disease and coronary heart disease (Horne, Camp, Muhlestein, & Cannon-Albright, 2006; Kauwe, Ridge, Foster, & Cannon-Albright, 2013).

Recently, Van den Berg et al. (2019) have proposed a simple familiarity rule. Using data from two sources, UPDB and LINKS (the Netherlands), they provide strong evidence that longevity is transmitted as a quantitative genetic trait among survivors up to the top 10% of their birth cohort. They also show that if you have first and second-degree relatives who survive to the top 10%, even if your own parents are not longevous, then you enjoy a survival advantage.

A large consortium of universities comprises the Long Life Family Study (LLFS) where they have developed the Family Longevity Selection Score (FLoSS). This score measures familiarity of longevity and was used to select families for LLFS but had not been validated in other populations. The LLFS team computed FLoSS using the lifespan data of 234,155 individuals from UPDB, born between 1779 and 1910 with mortality follow-up through 2012–2013. They reported (Arbeeva et al., 2018) that in UPDB those born after 1900 from 'exceptional' sibships based on the FLoSS had survival curves similar to that of the US participants from comparable LLFS probands. In this way, UPDB served as a basis from which others can test their methods of detecting familiarity of mortality. This study validated the FLoSS as selection criteria in family longevity studies using UPDB.

Clustering of mortality within families also aggregate with other important life history traits including fertility, in particular late female fertility. Women giving birth at advanced reproductive ages in natural fertility conditions have been found to have superior post-menopausal longevity (Smith, Mineau, & Bean, 2002). To determine if survival also improved for relatives of late-fertile women, Smith and colleagues (2009) compared male survival past age 50 for those with and without a late-fertile sister in two populations: UPDB (born 1800–1869) and the Programme de recherche en démographie historique in Québec (born 1670–1750). They reported improved male survival for those with, rather than without, a sister reproducing after age 45, suggesting that late female fertility and slower rates of aging may be promoted by similar genes. This work demonstrates again how UPDB promotes comparative work with other populations and in different eras.

The use of summary measures of familiarity is common using UPDB but UPDB has also advanced the application of statistical models that exploit the genealogies in UPDB but without summary estimates like FSIR or GIF. Garibotti and collaborators (2006) used pedigrees to assess the effects of unobserved environmental and genetic effects on longevity or so-called frailty. With UPDB they used two different frailty models that account for common environments (shared frailty) and genetic effects (correlated frailty). In a model that includes summary measures of familial history of longevity and both types of frailty effects, they found that genetic factors were comparable in their effects to shared environments.

## 3.2 INFANT MORTALITY

The detailed coverage of the population embodied within the UPDB facilitates the study of entire lives, from grávida to grave. It is equally important to investigate specific, primarily susceptible ages where mortality is highest: for infants and for those past mid-life. Lynch et al demonstrated during the early years of the UPDB, the power of genealogies for the study of infant mortality (Lynch, Mineau, & Anderton, 1985). They argue that the demographic study of declines in infant mortality during the last half of the 19th century was hampered by limited data at the individual-level; the US vital record system was not truly national until the 1930s. Accordingly, regional data served as the basis for the study of infant death. Here, they advocated for a form of family reconstitution drawn from the then-named Mormon Historical Project. This idea of using regional data as represented by the now-named UPDB was persuasive since it could be done using individual-level genealogy data. Moreover, the sweep of historical time covered by the UPDB was extensive and covered key and theoretically critical stretches of history encompassing the demographic transition beginning with the migration to Utah, an early natural fertility regime followed by the declining infant mortality and fertility rates. In addressing the value of these data, they report no pattern of sex or geography bias within the genealogies. Their study of early deaths showed that the slower decline in rural infant mortality was due to ongoing problems these areas had with access to health care relative to the more urbanized areas.

Bean, Mineau and Anderton (1992) showed how the timing and spacing of births altered the risks of infant death. Using an early Utah settlement cohort, they demonstrated that children born to older mothers with larger families and shorter birth intervals were more susceptible to 'contagion-and-competition effects'. They showed that during the early settlement of Utah, high levels of fertility were associated with early age at marriage, early childbearing, short birth intervals, and late ages at last birth. These fertility patterns are observed in a range of European and American populations before the quick secular decline in fertility and in many developing nations as outlined in the book *Fertility Change on the American Frontier* (Bean,

Mineau, & Anderton, 1990), populations that also have relatively high rates of infant mortality. Bean and colleagues also used UPDB to show how public health and medical practices served to alter the patterns of infant mortality including public health campaigns, use of herbs, improvement in water quality, and broader sanitation measures (Bean, Smith, Mineau, Fraser, & Lane, 2002).

### 3.3 SEX DIFFERENCES

The pattern of excess male mortality across the life span is well-known. What is also well-known is a paradox — women seem to be sicker while living longer and men seemingly healthier with shorter lifespans. In Utah, the prevalence of unhealthy male risk behaviors is lower than in most other male populations (due to lifestyle such as the prohibition on the use of tobacco and alcohol), whereas women experience greater mortality risks because of elevated fertility rates. UPDB was used by Lindahl-Jacobsen and his colleagues (2013) to assess how Utah's sex differential in mortality differs from Sweden and Denmark, given Utah has lower (in relation to the European nations) risk behaviors for males and higher fertility related risks for females, at least in the early settlement years. This prediction was not supported since Utah male-female differences in mortality were similar to that of Sweden and Denmark, suggesting a central role of biological mechanisms.

As the Lindahl-Jacobsen study suggests, male mortality may be exceeded by female mortality during the reproductive years. Penn and Smith (2007) showed that parents during 19th century Utah incurred fitness costs (i.e., excess mortality) from reproduction with women having higher mortality risks than men. They examined the survivorship and reproductive success of over 20,000 couples married between 1860–1895 and found that parity was negatively associated with parental survivorship, and more so for mothers than fathers. Increasing family size was also associated with lower offspring survival, primarily for later-born children, indicating a tradeoff between offspring quantity versus quality.

Additional analyses of UPDB by Bolund and her collaborators (2016) adopted a life-history perspective which predicts that reduced reproduction should benefit female lifespan when females endure greater costs of reproduction. They show a shift from male-advantaged to female-advantaged adult survival in individuals born before versus during the demographic transition. As fertility decreased over time, female lifespan increased, while male lifespan was stable, supporting the theory that differential costs of reproduction in the two sexes result in the shifting patterns of sex differences in lifespan across human populations.

Harrell and colleagues (2008) examined a novel question about sex differentials: are girls good and boys bad for parental longevity? They find significant but small adverse mortality effects for mothers after age 50 who bore mostly sons. Offspring sex composition did not have a significant effect on paternal mortality. Bearing mostly boys was found to be detrimental to maternal mortality regardless of childhood survival. This study is another example of the intergenerational value of the UPDB and how early circumstances (sex composition of offspring) alters mortality risks of parents.

### 3.4 FERTILITY AND MORTALITY TRADEOFFS

UPDB has been the basis for answering a central question in biology, anthropology, and biodemography: How are mortality risks associated with fertility patterns? One of the oldest questions in this domain relates to the association between consanguinity and survival. Jorde (2001) examined how parental consanguinity was associated with offspring mortality risks. To do this, he estimated inbreeding coefficients for over 300,000 Utahns born between 1847–1945 where approximately 3,500 inbred offspring were identified. For this analysis, elevated relative risks of pre-reproductive mortality were found among the offspring of first-cousin marriages and among the offspring of closer unions. Jorde argues that these mortality risks are larger in populations with low inbreeding and low mortality which allow one to see the effects of consanguinity more readily.

Smith and his colleagues (2002) considered the effects of fertility on longevity among mothers and fathers after age 60. They drew on evolutionary theories of aging and theories predicting social benefits and costs of children to older parents. Using UPDB data on 13,987 couples married between 1860–1899, they found that women with lower parity and those bearing children late in life lived longer post-reproductive lives. Husbands' longevity was less sensitive to reproductive history, although they faced mortality effects similar to their wives for more recent marriage cohorts. The fact that late age fertility during a natural fertility period was associated with better survival, especially for females, is consistent with the idea that slow reproductive senescence (e.g., late age at natural menopause as indicated by later fertility) is associated with overall somatic senescence. They find support for predictions based on evolutionary hypotheses about the tradeoffs

between fertility and mortality. As noted previously, the tradeoffs between fertility and longevity were replicated by Gagnon and Smith and their collaborators (2009) using three frontier populations.

## 3.5 SPECIAL POPULATIONS: WIDOWHOOD AND RELIGION

### 3.5.1 WIDOWHOOD

Mineau, Smith and Bean (2002) examined whether recently widowed individuals, male or female, have higher rates of mortality than comparable married persons over historical time. They employed life course analyses of four marriage cohorts extending from 1860 through 1904 with mortality follow-up to 1990. They found significant differences in the mortality risk for widowed men and women, with widowed men having excess mortality risks in every cohort and nearly every age. A consistent pattern of excess mortality in the comparison of married and widowed women was not observed. A key finding is that the relative mortality risks of widowhood, when they occur, have grown over time as secular trends of mortality decline.

Barclay and his colleagues (2020) addressed a little understood aspect of widowhood — how marital bereavement affects adult mortality in the context of polygamy. They studied over 200,000 men and women born before 1900 and their mortality into the 20th century. They showed that the death of a polygamist husband and the death of a 'sister' wife have similar adverse effects on female mortality. For men, the death of one wife in a polygamous marriage increases mortality to a lesser extent than it does for men in monogamous marriages. For polygamous men, losing additional wives has a dose-response effect. They also demonstrated that the presence of other kin in the household (a second wife, a sister wife, or children) attenuates the adverse effects of bereavement.

### 3.5.2 RELIGIOUS GROUPS

Questions regarding religion are not uncommon when using the UPDB. Mineau, Smith and Bean (2004) sought to examine how religious affiliation affects mortality risk. They examined all-cause mortality for a set of married men and women who survived to age 40 from selected birth cohorts (1850–1919). They found that individuals active in the Church of Jesus Christ of Latter-day Saints have lower mortality risks than those who are inactive or non-LDS in all cohorts and this relationship remains after controlling for socioeconomic status. The protective influences of being an active member of the LDS Church are greatest for the middle-aged and for those born in the more recent birth cohort. These results show that religious affiliation has stronger effects on adult mortality for men rather than women. These observations are consistent with explanations of health practices and social support factors that have been posited to understand the positive relationship between religious involvement and mortality outcome.

## 3.6 MORTALITY SUMMARY

UPDB offers considerable opportunities for research on all-cause and age-specific mortality for all ages and organized into families and pedigrees. The family-oriented nature of UPDB, with its deep genealogical information, has promoted from the very beginning the study of both historical family demographic analysis of mortality risks but also the potential role of genetic factors. We have highlighted how demography and genetics have provided synergistic guidance in the collection and use of mortality data. It is noteworthy that death certification begins in 1904 in Utah and these data, linked to UPDB, provide the very first UPDB 'medical' data elements in terms of cause of death. That this cause-specific information is now available and coded into the International Classification of Disease (ICD) schema has launched numerous studies and is another facilitating component of UPDB leading to collaborations between medical and demographic scholars. It is noteworthy that with linkages to contemporary electronic medical records, the morbidity profiles of individuals and how they may relate to the risk of death are now being analyzed extensively. Finally, the fact that families in Utah tend to be large (by US standards) and Utah is home to a large percentage of residents who are members of the Church of Jesus Christ of Latter-day Saints, these aspects of Utah's population as represented in UPDB attract the attention of researchers. Specifically, UPDB is used to examine the role of religion and family structure on longevity and how shifts in these two forces may be driving changes in mortality risks.

## 4 LIFE COURSE ANALYSIS OF THE EFFECTS OF EARLY LIFE CONDITIONS

The value provided by UPDB and other reconstructed historical databases arises because it permits life course analysis of data spanning full lifetimes and across generations, including genetics and shared traits across generations. UPDB especially allows for extensive and innovative assessment of early and mid-life conditions, including the presence of kin such as grandmothers, and their effects on later-life health and survival. The components of UPDB permit researchers to increase their ability to examine these associations with much greater precision for a large range of important early life factors and key later-life health outcomes. For example, UPDB contains hundreds of thousands of members of birth cohorts from the first half of the 20th century individuals for whom early and midlife conditions are measured and who are linked to their adult medical and mortality records generated decades later.

There are several advantages of UPDB for life course analyses. The complex data links in UPDB provide unparalleled data quality and depth, especially those that focus on families (nuclear, multigenerational, full pedigrees) and health outcomes that span entire life spans of individuals and their relatives. Given the data linkage model of UPDB, one can also study health effects of early life conditions and whether they generate direct effects on subsequent health outcomes or whether they operate through or are moderated by characteristics and circumstances arising during the adult years (e.g., widowhood, proximity to adult children). UPDB often provides ample data to study even moderate statistical interactions with sufficient power which generally require large sample sizes. Relatedly, the great advantage of data linked to the UPDB include key measures that are repeated over time allowing for the construction of trajectories that describe the dynamics of early life circumstances and later outcomes. Given its long-standing focus on genealogies, familial data in UPDB help to address confounding bias through the use of statistical models (fixed and random effects). This means that factors for which we lack direct measures but which are shared by family members (e.g., common family-of-origin environmental exposures, shared genes) can be introduced into the multivariate models.

### 4.1 INTERGENERATIONAL ANALYSIS

#### 4.1.1 GENETICS AND INTERGENERATIONAL ASSOCIATIONS

The influence of the UPDB on advancing our understanding of genetics and inheritance is legendary. Accordingly, it is not feasible to summarize the vast volume of genetic studies and discoveries in this essay. A few important highlights are described here that should be of interest to historical demographers. In 2004, *The New York Times* noted the value of UPDB and how Utah is proving to be an ideal genetic laboratory given the connections between genealogical data and medical records as well as biospecimens. Over a decade later in 2017, *The Atlantic* also illustrated the value of UPDB for genetic discoveries. It is often said that more diseases with genetic origins have been discovered in Utah than at any other university, an achievement attainable to a large extent due to UPDB. An earlier review of family-based genetic studies from a decade ago also provides a history and summary of the value of UPDB in terms of methodologies used to quantify the heritable contribution to traits and to identify genes potentially responsible for these traits (Cannon Albright, 2008).

An example of early genetic work that used the UPDB was by McLellan and colleagues (1984) where they compared gene frequency data for Utah with the gene frequencies from a U.S. population, 13 European populations, and seven populations from three religious isolates. The gene frequencies in Utah were found to be similar to those of their northern European ancestors. This is explained by the large founding size of the pioneer population in Utah and high rates of gene flow. More isolated groups such as the Amish, Hutterites, and Mennonites revealed more divergence from their ancestral populations and each other, due in part to social isolation. In a similar way O'Brien and her co-authors (1994) also examined the genetic structure of the Utah population using UPDB data.

One of the most prominent genetic discoveries that used UPDB was for identifying genes associated with breast and ovarian cancer risk in women (Miki et al., 1994). The identification of BRCA1 mutations has facilitated early diagnosis of breast and ovarian cancer susceptibility in some individuals as well as a better understanding of breast cancer biology. Known mutation carriers of this gene have been used in conjunction with the genealogies within UPDB to impute the genetic status of ancestors who lived decades and centuries before testing was possible (Smith, Hanson, Mineau, & Buys, 2012). This made it possible for a historical demographic study of genetic influences on fertility when modern contraception was not available. There

are other well-known studies of the identification of cancer mutations using UPDB, for example, melanoma and colon cancer (Cannon-Albright, Kamb, & Skolnick, 1996; Neklason et al., 2008).

While not using genetic (sequence) data, UPDB has been used effectively to study historical and intergeneration patterns of demographic phenomena. Anderton and his co-authors (Anderton, Tsuya, Bean, & Mineau, 1987) evaluated the hypothesized relationship between the fertility behavior of mothers and that of their daughters and the role of cohort effects. They further evaluated how both cohort-specific intermediate fertility determinants and mother's relative fertility behavior may explain specific fertility-timing patterns of daughters. Their analyses indicated that both fertility behavior and indirect associations regarding timing of fertility-related life-course events (e.g., marriage) are transmitted intergenerationally, that cohort-specific influences are substantial, and that intergenerational relationships may be more readily elaborated through the examination of fertility relative to cohort levels. Several years later Jennings and her colleagues (Jennings, Sullivan, & Hacker, 2012) used UPDB to show that during the onset of the fertility transition, reproductive behavior was transmitted across generations between women and their mothers, as well as between women and their husbands' family of origin. The findings suggest that the practice of parity-dependent marital fertility control and inter-birth spacing behavior derived, in part, from the previous generation and that the potential for mothers and mothers-in-law to help in the rearing of children encouraged higher marital fertility.

#### 4.1.2 GRANDMOTHER HYPOTHESIS

Given the generational depth of the UPDB, it has attracted attention by scientists interested in the role of social networks and family dynamics. A specific and influential topic has been the Grandmother Hypothesis. This argues that the long human female post-menopausal life span can be explained by the idea that grandmothers provide care to their grandchildren thereby increasing their fitness — facilitating the reproduction of their offspring and survival of grandchildren. A few exemplary papers demonstrate the impact and value of UPDB in addressing this question.

Hawkes and Smith (2009) noted that grandmother effects can be measured in data sets that include births and deaths over several generations, while recognizing unmeasured covariates complicate the task. They examined two complications: cohort shifts in mortality and fertility, and maternal age at death. They show that longevity of grandmothers may actually be associated with fewer grandchildren even when grandmother effects are actually positive (i.e., increased fitness). They further explored this question to address why humans evolved greater longevity while continuing to end female fertility at about the same age as some of our closest relatives, the great apes. With the grandmother hypothesis in mind, they compared age-specific mortality and fertility rates between humans and chimpanzees. They used 19th century women from UPDB to represent non-contracepting humans, and compared their fertility by age with published records for wild chimpanzees. They found wide individual variation in age at last birth in both humans and chimpanzees. This heterogeneity, combined with differences in adult mortality, has large and opposing effects on fertility schedules. There was support for the hypothesis that ages at last birth changed little while greater longevity evolved in humans.

To remain on the topic of the Grandmother Hypothesis but with focus on fertility, Dillon and collaborators (2020) assessed the role of grandmothers in fertility outcomes in a comparative historical demographic study based on four populations from Scandinavia (Sweden) and North America (two in Québec and Utah). The individual-level data, including UPDB, are all longitudinal and multigenerational, allowing them to address the impact of maternal and paternal grandmothers on the fertility of their daughters and daughters-in-law, while attending to heterogeneous effects across space and time as well as within-family differences via the use of fixed effects models. They found associations of paternal grandmother presence with higher fertility across the regions, as well as a general fertility advantage associated with the post-reproductive availability of the maternal grandmother. Overall, grandmothers were generally associated with high-fertility outcomes, but that the mechanism for this effect was co-determined by family configurations, resource allocation and the advent of fertility control.

## 4.2 INITIAL ANALYSIS OF EARLY LIFE CONDITIONS

For life course analyses, one of the first comprehensive assessments of childhood and young adulthood life and its effects on later life outcomes with UPDB was conducted by Smith and colleagues (2009). They considered how key early family circumstances affect mortality risks decades later. Early-life conditions were measured by parental mortality, parental fertility, religious upbringing, and parental SES. They noted an important issue: prior to these early-life conditions are familial and genetic factors that affect life span. Accordingly, they

examined the role of parental and familial patterns of longevity on mortality risks demonstrating the power of familial data by using frailty models to control for unobserved heterogeneity within families, all based on sib-pair data for 12,000 sib-pairs. They reported modest but significant effects of key childhood conditions (birth order, sibship size, parental religious affiliation, parental SES, and parental death in childhood). The effects of familial patterns of longevity were large and suggest that family history of key demographic measures may be an important but overlooked early life condition.

### 4.3 EARLY ADVERSITY AND LATER LIFE OUTCOMES: THE CASE OF FAMILY DEATHS IN CHILDHOOD

The range of early life conditions are manifold. For contemporary settings where individuals are asked to recall circumstances of their childhood and youth, batteries of questions exist such as the Adverse Childhood Experiences (ACE) instrument (Felitti et al., 1998; Greeson et al., 2014). Unfortunately, for historical demographers, the types of experiences and trauma encompassed in such measures are not generally available in the historic record. But this circumstance does not preclude demographers from exploring indicators of such events if they can be measured. The UPDB has identified a class of traumatic and adverse events that can be detected: deaths of family members when a person is young. This class of variables is arguably an unambiguous indicator of stress in childhood, is visible in most historical databases and when family relationships are well measured, and one can analyze how the type of death (sibling, parent, child) and cause of death may have differing effects.

#### 4.3.1 SIBLINGS AND OFFSPRING

Van Dijk, Janssens and Smith (2019) observed that the literature has yielded mixed evidence about the influence of infant and child mortality in *birth cohorts* on adult mortality. These studies generally do not examine the specific role of mortality *within a family context* when micro data are available. They examined how exposure to mortality as a child is related to their adult mortality risk between ages 18 and 85 in UPDB (1874–2015) and the LINKS data from Zeeland (the Netherlands) 1812–1957. They found that childhood exposure to community mortality and sibling deaths increases adult mortality rates. Effects of sibling mortality on adult all-cause mortality risk were stronger in Utah, where sibling deaths were less common in relation to Zeeland. Exposure to sibling deaths from infection was related to the siblings' risk of adult mortality from cardiovascular disease and diabetes mellitus, a result consistent with an inflammatory immune response mechanism.

The direct measure of biomarkers for inflammation were available in a separate study (based on the Cache County Memory and Health Study linked to UPDB) which showed that sibling deaths also elevated inflammation, a key biomarker for mortality, as measured by high-sensitivity C-reactive protein (CRP) (Norton, Hatch, Munger, & Smith, 2017). This study demonstrated a link between significant psychosocial stress in early life and immune-inflammatory functioning in late life, and reinforces a mechanism explaining the link between early-life adversity and late-life health.

#### 4.3.2 PARENTS: EFFECTS ON OFFSPRING MORTALITY

A study by Smith and colleagues (2014) asked whether a parental death is associated with enduring mortality risks after age 65? The years following parental death may initiate circumstances in which the adverse effects of paternal death operate. Accordingly, they examined the offspring's marital status, adult SES, fertility, and later-life health status, where the latter relies on the comprehensive Charlson Comorbidity Index (Charlson, Szatrowski, Peterson, & Gold, 1994) using Medicare data. They show that offspring whose parents died when they were children, but especially when they were adolescents/young adults, have modest but significant mortality risks after age 65. Strikingly, there were weak mediating influences of later-life comorbidities, marital status, fertility and adult socioeconomic status.

One of the hypothesized effects of parental death is suicide risk of the offspring. Hollingshaus and Smith (2015) examined this question while also considering whether the surviving parent remarried. Using UPDB for birth cohorts between 1886 and 1960 (N = 663,729, including 4,533 suicides), they demonstrated that parental death was associated with an excess risk of adult offspring suicide before age 50, and with increased risk of cardiovascular disease deaths (CVD) for adults of all ages. Daughters whose surviving parents remarried had (in relation to those who did not remarry) a smaller risk of suicide before age 50 (though not statistically significant), but significantly higher risk after age 50. Parental remarriage had no effect on male suicide risk. This analysis illustrates the value of using death certificate and detailed health information as a method to explicate possible mechanisms linking early events to later health outcomes.

### 4.3.3 PARENTS: EFFECTS ON OFFSPRING ALZHEIMER'S DISEASE (AD) RISK

With UPDB, all-cause mortality and suicide specifically have been identified as risks for those experiencing the death of a parent. This is possible with UPDB and other databases linked to vital records. UPDB has also had impact by virtue of links to cohorts whose clinical assessments and questionnaire response have been linked to UPDB. This has led to a series of papers on parental death in childhood and AD risk.

In a pair of papers, Norton and colleagues (2009, 2011) examined early parental death and late-life dementia risk in offspring based on links between UPDB and the Cache County Memory and Health Study. They showed parental death during one's childhood is associated with higher prevalence of AD, with different effects based on the ages of an individual when they experienced father's versus mother's death. The strength of these associations was attenuated by remarriage of the widowed parent.

The same team of investigators examined family deaths more broadly using UPDB genealogical and mortality data and their effects on AD risk (Greene et al., 2014; Norton et al., 2016). Norton and her team examined whether parental death during one's childhood, and offspring and spouse deaths during adulthood are associated with faster cognitive decline and higher Alzheimer's disease (AD) risk in late life. Using 4,545 non-demented participants from the Cache County Memory and Health Study linked to UPDB found that age moderated the relationship between family deaths and AD. For persons aged 65–69 years at baseline those exposed to more deaths during adulthood faced a two-fold AD risk whereas those over 80 years had a lower AD risk. Their findings again emphasized the value of linking family history from UPDB to outcomes from an epidemiological cohort where they were able to demonstrate an effect about the link between family member deaths during adulthood and AD risk later in life. In a related study, Greene and his collaborators (Greene et al., 2014) tested the hypothesis that experiencing an offspring death was associated with an increased rate of cognitive decline in late life. They also used UPDB linked to the Cache County Memory and Health Study based on 3,174 non-demented residents aged 65–105. They reported that subjects who experienced offspring death before age 30 experienced a significantly faster rate of cognitive decline in late life, but only if they had an *APOE e4* allele (a major genetic risk factor for Alzheimer's Disease). Also, an offspring death was only related to faster cognitive decline when there were no subsequent births. Other stressful life events were also shown to affect the rate of cognitive decline in this cohort (Tschanz et al., 2013).

## 4.4 LIFE COURSE SUMMARY

The depth of familial data spanning over 200 years contained with UPDB creates an opportunity to link the circumstances of an individual early in their lives to events over all years of their existence until death or at least for many decades. With the genealogical structure of UPDB, the data can then link the circumstances of the parents and more distant ancestors to the life course of the target individual for fertility, health, socioeconomic, and mortality outcomes. UPDB data provide the kind of data that supports and is consistent with the perspective advanced by the Developmental Origins of Health and Disease (DOHaD) concept (Gage, Munafò, & Davey Smith, 2016; Hanson, Poston, & Gluckman, 2019; Penkler, Hanson, Biesma, & Muller, 2019). Overall, deep longitudinal data with nearly complete family ascertainment over many generations provide the necessary ingredients for life-course research. Indeed, in the contemporary period, it is now possible to consider the role of 'shocks' long ago and how they may affect the expression of genetic predispositions (i.e., epigenetics) today. It is also worth noting that the challenge for life course research is the vexing problem of loss to follow-up that is unavoidable when so many years of follow-up for millions of families are involved. UPDB takes steps to monitor and time-stamp, when possible, the arrival to and departure from Utah.

## 5 SPECIAL TOPICS

### 5.1 SOCIO-ECONOMIC STATUS (SES)

It is well known that socioeconomic status (SES) is a central facet affecting and being affected by demographic structure and change. Nonetheless, users of historic records seeking to identify measures of SES face challenges given the vagaries of the source records and their completeness and consistency over time. UPDB data are not immune to these challenges. Despite this, UPDB data contain relatively consistent measures of SES over many decades that rely most heavily on occupation and industry. These data are derived primarily from vital

records (birth and death certificates; birth from 1915 and death from 1904) and from the US Census of Utah for the available decennial censuses spanning 1880–1940. These data pertain to individuals (adults and by extension, to their offspring). Given the depth of genealogical and spatial information in UPDB, it is possible to construct SES measures of kin to determine a familial composite SES indicator of a person's kin and their geographic proximity. 'Neighborhood' or community indicators of SES can also be derived within UPDB data as well through the more conventional methods of linking location data in UPDB data to Census (or any other geo-referenced) records at the county, census tract or census block group level.

A persistent question exists in the study of SES and health outcomes, particularly mortality risk: has the inverse relationship between SES and mortality risk always existed or is it a recent phenomenon? Bengtsson, Dribe and Helgertz (2020) have acknowledged that today there is a consistent mortality pattern by SES at all ages but not all confirm this association. They note that if a gradient did not exist in the past, then when did the association begin. They conclude that adult mortality risks for men and women in southern Sweden over a 200-year period was associated with social class risks emerging for middle-aged persons only after 1950 for women and after 1970 for men, and later for ages 60–89. These findings occurred when Sweden became a modern welfare state with universal health care system, suggesting the importance of psychosocial factors. In contrast, Smith and his team (2009) examined with UPDB data how key early family circumstances affect mortality risks decades later including parental mortality, parental fertility religious upbringing (Mineau et al., 2004), and *parental SES* for individuals born in the 19th century. Using frailty models for 12,000 sib-pairs, they found significant adult mortality effects associated with parental SES in childhood. This suggests that during a period of limited formal social support mechanisms, an inverse SES-mortality association was detected in Utah.

The conduct of innovative studies using SES in UPDB is represented in several ways. Temby and Smith (2014) elaborated on the SES and mortality association by considering the interaction between SES and a positive family history of longevity. They considered, for example, whether individuals with lower levels of SES may experience an attenuated longevity penalty if they have long-lived relatives. Analysis of survival past age 40 for men born between 1840 to 1909 showed that mortality risks for men with the highest SES was reduced more as familial longevity increased than it does for the lowest SES men. Mortality risks for farmers also declines more as familial longevity increases in relation to non-farmers. They suggest that a type of gene-environment interaction occurs whereby the benefits of a family history of longevity are more available to those who have higher status occupations.

A novel examination between the interplay of the SES of offspring and the mortality risks of their parents was considered by Zimmer and his colleagues (2016b). They tested the hypothesis that SES effects may 'flow up' from offspring to parents: higher offspring SES associates with lower parental mortality after age 40 after controlling for parental SES. They used 30,000 individuals born between 1864–1883 whose offspring were born between 1886–1920 where SES was based on the Nam-Powers occupational status scores (Nam & Boyd, 2004; Nam & Powers, 1983) divided into quartiles and a category for farmers. They showed a longevity penalty for parents whose offspring have low SES and a longevity dividend for those with high-SES offspring. They expanded on this hypothesis (Zimmer, Hanson, & Smith, 2016a) by considering morbidity as measured by the Charlson Comorbidity Index (Charlson et al., 1994). They used sex-specific group-based trajectory patterns of morbidity and survival where group morbidity trajectories were ranked from least to most healthy. They showed that increasing (one's own) SES in childhood is associated with membership in groups that have more favorable morbidity trajectories as well as survival probabilities. SES in adulthood has additive impact, especially for females. These two studies illustrate the insights gained from both longitudinal and intergenerational analytic strategies: the influence of offspring SES on well-being of parents and the role of both childhood and adult SES in independently influencing old-age morbidity risks.

UPDB data have been effective in helping to expose SES differentials in fertility as well. Dribe and his cross-national comparative team (2017) used longitudinal individual-level data from five populations in Europe and North America to examine linked SES and fertility during the fertility transition. They specifically studied the dynamism of SES differences in marital fertility and related these to fertility behavior during the demographic transition. They found no support for the hypothesis of universally high fertility among the upper classes in pre-transitional society, but did provide findings consistent with the hypothesis that the upper classes preceded other groups in reducing their fertility. Farmers and unskilled workers were the latest to start limiting their fertility. Within Utah, Maloney, Hanson and Smith (2014) used UPDB to examine differences across occupational classes in fertility levels and in the timing and pace of change in fertility in Utah in the late 19th and early 20th centuries. They showed that families of white-collar workers led changes in many fertility-related behaviors including age at first marriage and first birth interval while farm families continued

to have high fertility levels and bore children into later ages. They identified patterns of fertility change tied to variation in important economic circumstances such as the length of education and training required for particular occupations, or the need for family-based labor on the farm.

## 5.2 TWINS

Given the richness of the genealogical, and therefore fertility information of UPDB, a specific topic has attracted the attention of historical demographers: twins. We note that UPDB does not comprehensively contain zygosity information so it is not possible to assess identical and fraternal twins in the database. The exception relates to opposite sex twins who are by definition dizygotic (non-identical). One of the earliest examination of twins using UPDB was work by Carmelli, Hasstedt and Andersen (1981). They investigated demographic and genetic aspects of human twinning. They found that Utah has an elevated incidence of twinning in relation to the US white population. They concluded that most of the general decrease in twinning during the 19th and 20th centuries was due to a maternal age effect. Couples bearing children in later US birth cohorts (after 1900) began to practice family limitation, limiting fertility at older ages. Utah couples did not significantly alter their behavior during the early years of childbearing so their fertility remained high with an average number of children above four thereby enhancing the number of twin births.

Decades later, Robson and Smith (2011, 2012) tested two hypotheses regarding twinning in human populations that have alternative predictions about the effects of bearing twins on maternal life time reproduction and survival. The maternal depletion hypothesis argues that mothers of twins will suffer negative outcomes while a 'robustness' hypothesis argues that while twinning is costly, it may reveal mothers with a greater capacity to bear that cost. Using UPDB, they examine mothers who lived at least to the age of 50 and found evidence consistent with the robustness hypothesis: mothers of twins had lower postmenopausal mortality, shorter average inter-birth intervals, later ages at last birth and higher lifetime fertility than their singleton-only bearing counterparts. They concluded that bearing twins is more likely for those with a robust phenotype and may be a useful indicator of maternal heterogeneity.

While the robustness hypothesis was supported with respect to the mother, a follow-up study (Chernenko, Hollingshaus, Robson, Hanson, & Smith, 2018) examined mortality patterns for the singleton offspring of mothers who had twins compared to the single offspring of mothers who had not had twins to determine whether they share the hypothesized robust phenotype of their mothers. They showed that singleton offspring of twinning mothers experience a survival disadvantage prior to age 5, no survival benefit or penalty between ages 5 and 49, and — for males only — a significant survival advantage after age 50. They also found a survival disadvantage in early life for singleton offspring of twinning mothers born immediately after the twinset for both sexes. They conclude that while bearing twins may reflect a robust maternal phenotype, the toll of bearing twins may disadvantage subsequent offspring, especially during infancy.

## 5.3 SEX RATIOS

Investigators have used UPDB to examine several factors associated with a key structural demographic measure, sex ratios, and their consequences. Analysis of the fertility histories of women born between 1850–1900 by Bohnert and colleagues (2012) considered whether there was evidence of sex preference in these early decades of Utah's history. They found more male children, as expressed by birth stopping behavior after the birth of a male child and shorter birth intervals in higher-parity births when most previous children were female. Evidence was presented focusing on two sub-populations, farmers and individuals with stronger ties to the Church of Jesus Christ of Latter-day Saints. They showed that the former, while having relatively high fertility rates, had similar preferences for male children as the other Utahns.

Farmers, who presumably had a need for family labor, were more interested in the quantity than in the sex mix of their children. Schacht and Smith (2017) examined the historical patterns of sex ratios and their importance in understating shifts in other demographic phenomena. They further evaluated whether the sex ratio at birth (SRB) may be patterned by maternal condition and/or environmental stressors (Schacht, Tharp, & Smith, 2019). Using UPDB data for the population during the interwar period (1918–1939), inclusive of three distinct eras (Spanish Flu, Roaring '20 s, and the Great Depression), they assessed two theoretical frameworks used to study patterning in SRB — (1) 'frail males' and (2) adaptive sex-biased investment theory (Trivers & Willard, 1973). The first approach centers on greater male susceptibility to exogenous stressors and argues that offspring survival should be expected to differ between 'good' and 'bad' times. The second approach predicts that mothers themselves play a direct role in manipulating offspring SRB, and that those in better condition should invest more in sons. Consistent with the 'frail male' predictions, they found that boys are less likely to be born during the environmentally challenging times of the Spanish Flu and Great

Depression. However, they found no evidence that maternal condition is associated with sex ratios at birth, a result inconsistent with the Trivers-Willard hypothesis.

## 5.4 SPECIAL TOPICS

Topics highlighted here are, quite simply, key selected examples of demographic interest but there are certainly many others. We note that other cross-cutting topics include studies that are spatially oriented (Smith et al., 2011; Stroup et al., 2017; Zick et al., 2009) as well as less-studied family formation topics such as those investigating step-children (Schacht, Meeks, Fraser, & Smith, 2021). The range of possible topics is considerable and in cases where the event is rare (e.g., extreme longevity, very young fertility) or is likely to vary by context (e.g., different centuries or nations), the size and breadth of UPDB lends itself to comparative analyses (e.g., Dillon et al., 2020; Dribe et al., 2017; Gagnon et al., 2009; Smith et al., 2009).

## 6 CONCLUSIONS AND FUTURE DIRECTIONS

Many historical databases represent considerable investments of time, talent, and institutional commitment. This characterization certainly fits the UPDB. While the ideas and actions of hundreds of investigators and their academic departments and universities have already generated considerable value to scientific advances broadly and demography specifically, the future provides us with additional opportunities and challenges.

### 6.1 GENETICS AND DEMOGRAPHY

We have described numerous instances where geneticists and demographers have benefitted from each other's perspectives and methodologies (Adams, Lam, Hermalin, & Smouse, 1990; Bean, 1990). While they inform each other, the UPDB and its deep genealogical data are based on high quality and desirably redundant data (i.e., multiple reports of the same family connections) which are not always confirmed genetically. At its core, between genealogies from the Genealogical Society of Utah and vital records as well as other sources, there is the possibility that some of the genetic relationships may need verification. Mistaken connections are rare but happen nonetheless due to the human nature of record keeping, regulations that protect identities (e.g., birth versus adoptive parents on birth certificates), and explicit actions to insulate the parties from adverse consequences (e.g., naming an individual as the father incorrectly). The rise of 'Big Genetics' (Smith, Hanson, & Mineau, 2016) is increasingly generating sequence data that is quantifying the likelihood that two persons are related genetically and in what way. Validating the genealogical connections this way, to the extent that these DNA measures are suitably available, may improve the quality of the UPDB and its genealogies. Similarly, genealogies may also provide corrections to the processing of genetic information, which also has imperfections, and serve to help correct erroneously linked persons identified through sequence data.

The burgeoning field of biodemography over recent decades, emerging against the backdrop of the Human Genome Project and the HapMap Project, reflects the union of ideas and data derived from demography, evolutionary biology, genetics, and medicine. Recognition of these connections was central to the concept and creation of the UPDB — and outlined in *Convergent Issues in Genetics and Demography* (Adams et al., 1990) in a chapter by Dr. Lee L. Bean, one of the original co-developers of the UPDB titled 'Utah Population Database: Demographic and Genetic Convergence and Divergence'. This early insight has led to several key genetic discoveries and insights (Adams et al., 1990; Cawthon et al., 2020; Cawthon, Smith, O'Brien, Sivatchenko, & Kerber, 2003; Hanson et al., 2020; Miki, et al., 1994; Neklason et al., 2008; Norton et al., 2010, 2016; Smith et al., 2012). The contribution of the UPDB to genetics is considerable and per se is beyond the scope of this review but that specific impact of the UPDB is acknowledged below.

We can imagine demographic analyses that can effectively control for genetic variants that may contribute to our understanding of central outcomes such as fertility and mortality. This appears to be more common in epidemiology (e.g., Carreras-Torres et al., 2014) where polygenic risk scores (summary values of overall risk as measured across a range of many genes and alleles) but it is easy to see how this may attract the interest of demographers.

One example noted previously may be applied more broadly which is to use known carriers of certain genetic variants in the population and then exploit information about family lineages and modes of inheritance to

identify (obligate) carriers in the past based on where a person is in the family pedigree. For example, if we know that two children have a specific genetic variant, and the disease follows an autosomal dominant mode of inheritance, we can assume that one of the parents (the common ancestor) is also an obligate carrier. This can be expanded to more distant relatives. This method was applied using BRCA1 mutations (which increase the risk of breast and ovarian cancer) in modern samples and studying mortality and fertility differentials of common ancestors for whom we can ascertain their mutation status (Smith et al., 2012).

For demographers who generally work with large and sometimes complete population data of individuals, they would prefer to have genetic information on all persons under investigation along with other commonly analyzed covariates. This is an ambition and is slowly becoming available in certain data sets such as the US Health and Retirement Survey but it is still rare, though the lower costs for genetic sequencing is making this approach more affordable. As genetic data become more available for large samples, the prospects for conducting gene x environmental interactions, obligate carriers analyses and social genomics (Das, 2021; Sanz-de-Galdeano, Terskaya, & Upegui, 2020) will become more common. For UPDB, it is the norm that selected families and pedigrees displaying traits of interest are recruited and their DNA samples collected, rather than wholesale genetic data collection of large sections of the population.

With UPDB, it is worth pointing out that in cases where genomic sequence data to characterize possible genetic predispositions are lacking on a population basis, it is quite feasible to use family history of a trait as an empirical and practical alternative. Use of such family history information has been described in this review and is used as a broader indicator of shared genes and environments. While admittedly a different type of 'genetic' risk indicator, it offers several advantages such as being feasible, inexpensive, adaptable to the type of inheritance pattern (e.g., autosomal dominant, maternal (mitochondrial) inheritance), and having the flexibility for investigators to create multiple family history measures for different traits or behaviors. Others have advocated the use of family history as an attractive tool for assessing possible genetic susceptibility (Rich et al., 2004; Scheuner, Wang, Raffel, Larabell, & Rotter, 1997).

## 6.2 GEOGRAPHY AND DEMOGRAPHY

These two disciplines have been close cousins for centuries and UPDB reflects this close connection and accordingly it contains vast data elements important to both. UPDB has devoted resources to enhance the ability of investigators to test spatially-oriented hypotheses linked to health and demographic outcomes. Geo-referencing spatial data in order to locate individuals and their relatives and neighbors over decades has been a focus of UPDB (Leiser et al., 2020; Stroup et al., 2017). The challenge is to link persons to an area with comparable precision subject to data constraints. In the US, ZIP codes may be available but not precise addresses. In the more distant past, place names are generally available and locating them with consistency with systematic coding is on-going and generally achievable, especially when the need is to study complete lives and all their locations. Again, inclusion of some persons for whom there is geographic information may come at the price of precision. Certainly, working with other organizations, notably the National Historical Geographic Information System (NHGIS) with IPUMS is providing invaluable benefits in this respect. The increasing availability of environmental exposure data, such as from the US Environmental Protection Agency's Air Quality System Data Mart, can be layered on top of geo-referenced data (precise spatial coordinates, US census tracts or block groups) to develop environmental exposures to support studies of their influence on demographic outcomes.

## 6.3 CENSUS RECORDS

UPDB contains linked individual data from the US Censuses of Utah from 1880–1940. This achievement has taken place over the past 40 years motivated and funded by specific research projects but the resulting Census data are now a feature of UPDB. Given the GIS data and geo-referenced persons at various points in time during their lives, it is also common to link areal data from the Census (and other GIS data sources) to UPDB. This historic geospatial information may have broad area indicators (Census Enumeration Districts identifiers) which may be used to control for shared environmental features without necessarily knowing what those precise features are.

The University of Utah now houses one of the Federal Statistical Research Data Centers (FSRDC), a facility that provides access to otherwise secure data collected by the federal government and made available to approved investigators. The FSRDC infrastructure allows other data, such as the UPDB, to link to the US Census and data it collects, if approval is secured from the Census and the institutional owners of the data contained within UPDB. This linkage provides an outstanding opportunity to link securely and with privacy

safeguards UPDB to other federal data such as the decennial Census and the American Community Survey. Such linked data can then be analyzed within the FSRDC (The Wasatch Front Research Data Center at the University of Utah).

## 6.4 ANCESTORS THEN, DESCENDANTS NOW

The UPDB grows every year as new data accrue to the agencies who collect them, who in turn provide approved data to UPDB for linkage. The newly installed data represent both additional information about a given individual as well as new descendants being added to a Utah family tree either through birth, marriage or migration. This allows new opportunities to examine how prior events or genetic signals continue to impart their influence on current residents of Utah.

As discussed in the section on Life Course Analysis, this in-depth and ever-increasing volume of allows demographers to think broadly about how the configuration of one's family pedigree and their spatial orientation may have previously unknown effects on key outcomes. Additional questions can be addressed related to the presence of a grandmother and daughter fertility, density of kin in the neighborhood and survival, adoption of reproductive behavior as a function of that behavior in close relatives, and the volume and nature of specific causes of death in relatives and how that alters an individual's risk of death from those causes.

## 6.5 BRIDGING DATABASES AND COMPARATIVE ANALYSES

The growth of historical data bases throughout the globe give rise to the potential of comparative research but also direct collaboration by identifying family lineages that touch more than database. In the case of UPDB, the founding European population was based in northern and western Europe. For example, Scandinavian and British founders served as the basis of the early pioneers who settled Utah's early frontier expansion. Accordingly, many present data Utahns can claim their heritage from those parts of Europe and indeed common ancestors appear in UPDB and several of the European historical demographic databases. The opportunity exists to examine the demographic impact of staying versus leaving one's home country — some left Europe to the US and others remained. What has become of these two broad segments of a family and their respective descendants? Indeed, what can we learn from those who left Europe for the US only to return later?

## 6.6 FINAL THOUGHTS

The UPDB has contributed to the growth and development of demography through the sheer number of people and families represented in its data holdings and the wide-ranging data available about each individual. But we recognize that UPDB has succeeded for other reasons. One factor for UPDB's success relates to the relatively small size of Utah's population and institutions. Utah's small population at the outset in the mid-1970s likely contributed to its launch. The volume of data was more manageable and the ability of key institutions to interact was conducive to creating a collaborative atmosphere between the initial participating institutions: The Genealogical Society of Utah, The University of Utah, and the Utah Department of Health. On the latter point, the geographic proximity of these institutions contributed to interactions, negotiations and agreements that would likely have been more problematic in much larger states. The consistent and robust support of the University of Utah and the Huntsman Cancer Institute to maintain and fund the UPDB is without question a key ingredient in explaining the success of the database.

In the end, the growth and evolution of investigators and topics reliant on UPDB can in part be attributed to the catalyzing effects of big data on team science (Sellers, Caporaso, Lapidus, Petersen, & Trent, 2006; Shah, Pico, & Freedman, 2016; Stokols, Misra, Moser, Hall, & Taylor, 2008). The diversity and quality of UPDB data that is curated and made safely available has served to induce many ambitious projects that involve investigators from multiple disciplines that would not have been possible otherwise. This has created teams that often combine medical, population and social sciences. These multidisciplinary efforts serve to strengthen the science under investigation.

## ACKNOWLEDGEMENTS

Support for this work was partially provided by R01 AG022095 (Early Life Conditions, Survival, and Health: A Pedigree-Based Population Study). We also thank the Pedigree and Population Resource of Huntsman Cancer Institute, University of Utah (funded in part by the Huntsman Cancer Foundation) for its role in the ongoing collection, maintenance and support of the Utah Population Database (UPDB). We also acknowledge partial support for the UPDB through grant P30 CA2014 from the National Cancer Institute, University of Utah and from the University of Utah's program in Personalized Health and Center for Clinical and Translational Science.

## REFERENCES

- Adams, J., Lam, D. A., Hermalin, A. I., & Smouse, P. E. (Eds.) (1990). *Convergent issues in genetics and demography*. New York: Oxford University Press.
- Anderton, D. L., & Bean, L. L. (1985). Birth spacing and fertility limitation: A behavioral analysis of a nineteenth century frontier population. *Demography*, 22(2), 169–183. doi: [10.2307/2061176](https://doi.org/10.2307/2061176)
- Anderton, D. L., Bean, L. L., Willigan, J. D., & Mineau, G. P. (1984). Adoption of fertility limitation in an American frontier population: An analysis and simulation of socio-religious subgroups. *Social Biology*, 31(1–2), 140–159. doi: [10.1080/19485565.1984.9988569](https://doi.org/10.1080/19485565.1984.9988569)
- Anderton, D. L., Tsuya, N. O., Bean, L. L., & Mineau, G. P. (1987). Intergenerational transmission of relative fertility and life course patterns. *Demography*, 24(4), 467–480. doi: [10.2307/2061386](https://doi.org/10.2307/2061386)
- Arbeeva, L. S., Hanson, H. A., Arbeev, K. G., Kulminski, A. M., Stallard, E., Ukraintseva, S. V., . . . Yashin, A. I. (2018). How well does the family longevity selection score work: A validation test using the Utah Population Database. *Front Public Health*, 6, 277. doi: [10.3389/fpubh.2018.00277](https://doi.org/10.3389/fpubh.2018.00277)
- Barclay, K. J., Thorén, R. D., Hanson, H. A., & Smith, K. R. (2020). The effects of marital status, fertility, and bereavement on adult mortality in polygamous and monogamous households: Evidence from the Utah Population Database. *Demography*, 57(6), 2169–2198. doi: [10.1007/s13524-020-00918-z](https://doi.org/10.1007/s13524-020-00918-z)
- Bean, L. L. (1990). The Utah Population Database: Demographic and genetic convergence and divergence. In J. Adams, D. A. Lam, A. I. Hermalin & P. E. Smouse (Eds.), *Convergent issues in genetics demography* (pp. 231–244). New York: Oxford University Press.
- Bean, L. L., & Mineau, G. P. (1986). The polygyny-fertility hypothesis: A re-evaluation. *Population Studies*, 40(1), 67–81. doi: [10.1080/0032472031000141846](https://doi.org/10.1080/0032472031000141846)
- Bean, L. L., Mineau, G. P., & Anderton, D. L. (1990). *Fertility change on the American frontier: Adaptation and innovation*. Berkeley: University of California Press.
- Bean, L. L., Mineau, G. P., & Anderton, D. L. (1992). High-risk childbearing: Fertility and infant mortality on the American frontier. *Social Science History*, 16(3), 337–363. doi: [10.1017/S0145553200016539](https://doi.org/10.1017/S0145553200016539)
- Bean, L. L., Smith, K. R., Mineau, G. P., Fraser, A., & Lane, D. (2002). Infant deaths in Utah, 1850–1939. *Utah Historical Quarterly*, 70(2), 158–173.
- Bengtsson, T., Dribe, M., & Helgertz, J. (2020). When did the health gradient emerge? Social class and adult mortality in southern Sweden, 1813–2015. *Demography*, 57(3), 953–977. doi: [10.1007/s13524-020-00877-5](https://doi.org/10.1007/s13524-020-00877-5)
- Bohnert, N., Jåstad, H. L., Vechbanyongratana, J., & Walhout, E. (2012). Offspring sex preference in frontier America. *The Journal of Interdisciplinary History*, 42(4), 519–541. doi: [10.1162/JINH\\_a\\_00303](https://doi.org/10.1162/JINH_a_00303)
- Bolund, E., Lummaa, V., Smith, K. R., Hanson, H. A., & Maklakov, A. A. (2016). Reduced costs of reproduction in females mediate a shift from a male-biased to a female-biased lifespan in humans. *Scientific reports*, 6, 24672. doi: [10.1038/SREP24672](https://doi.org/10.1038/SREP24672)
- Cannon Albright, L. A. (2008). Utah family-based analysis: Past, present and future. *Human Heredity*, 65(4), 209–220. doi: [10.1159/000112368](https://doi.org/10.1159/000112368)
- Cannon-Albright, L. A., Kamb, A., & Skolnick, M. (1996). A review of inherited predisposition to melanoma. *Seminars in Oncology*, 23(6), 667–672.
- Carmelli, D., Hasstedt, S., & Andersen, S. (1981). Demography and genetics of human twinning in the Utah Mormon Genealogy. *Progress in Clinical and Biological Research*, 69A, 81–93.
- Carreras-Torres, R., Kundu, S., Zanetti, D., Esteban, E., Via, M., & Moral, P. (2014). Genetic risk score of NOS gene variants associated with myocardial infarction correlates with coronary incidence across Europe. *PLoS ONE*, 9(5), e96504. doi: [10.1371/journal.pone.0096504](https://doi.org/10.1371/journal.pone.0096504)

- Casey, J. A., Schwartz, B. S., Stewart, W. F., & Adler, N. E. (2016). Using electronic health records for population health research: A review of methods and applications. *Annual Review of Public Health, 37*, 61–81. doi: [10.1146/annurev-publhealth-032315-021353](https://doi.org/10.1146/annurev-publhealth-032315-021353)
- Cawthon, R. M., Meeks, H. D., Sasani, T. A., Smith, K. R., Kerber, R. A., O'Brien, E., . . . Jorde, L. B. (2020). Germline mutation rates in young adults predict longevity and reproductive lifespan. *Scientific reports, 10*(1), 10001. doi: [10.1038/s41598-020-66867-0](https://doi.org/10.1038/s41598-020-66867-0)
- Cawthon, R. M., Smith, K. R., O'Brien, E., Sivatchenko, A., & Kerber, R. A. (2003). Association between telomere length in blood and mortality in people aged 60 years or older. *Lancet, 361*(9355), 393–395. doi: [10.1016/S0140-6736\(03\)12384-7](https://doi.org/10.1016/S0140-6736(03)12384-7)
- Charlson, M., Szatrowski, T. P., Peterson, J., & Gold, J. (1994). Validation of a combined comorbidity index. *Journal of Clinical Epidemiology, 47*(11), 1245–1251. doi: [10.1016/0895-4356\(94\)90129-5](https://doi.org/10.1016/0895-4356(94)90129-5)
- Chernenko, A., Hollingshaus, M., Robson, S., Hanson, H. A., & Smith, K. R. (2018). Tykes, toddlers, teens, and twins of robust mothers: Do the offspring of twinning mothers share in their mother's robust phenotype? *Biodemography and Social Biology, 64*(2), 102–113. doi: [10.1080/19485565.2018.1486697](https://doi.org/10.1080/19485565.2018.1486697)
- Chernenko, A., Meeks, H., & Smith, K. R. (2019). Examining validity of body mass index calculated using height and weight data from the US driver license. *BMC Public Health, 19*(1), 100. doi: [10.1186/s12889-019-6391-3](https://doi.org/10.1186/s12889-019-6391-3)
- Das, A. (2021). Transpersonal genetic effects among older U.S. couples: A longitudinal study. *The Journals of Gerontology: Series B, 76*(1), 184–194. doi: [10.1093/geronb/gbz151](https://doi.org/10.1093/geronb/gbz151)
- Dillon, L., Chernenko, A., Dribe, M., Engelhardt, S., Gagnon, A., Hanson, H. A., . . . Vézina, H. (2020). Did grandmothers enhance reproductive success in historic populations?: Testing evolutionary theories on historical demographic data in Scandinavia and North America. In O. Burger, R. Lee & R. Sear (Eds.), *Human Evolutionary Demography*. Retrieved from <https://osf.io/qgjp8/>
- Dribe, M., Breschi, M., Gagnon, A., Gauvreau, D., Hanson, H. A., Maloney, T. N., . . . Vézina, H. (2017). Socio-economic status and fertility decline: Insights from historical transitions in Europe and North America. *Population Studies, 71*(1), 3–21. doi: [10.1080/00324728.2016.1253857](https://doi.org/10.1080/00324728.2016.1253857)
- DuVall, S. L., Fraser, A. M., Rowe, K., Thomas, A., & Mineau, G. P. (2012). Evaluation of record linkage between a large healthcare provider and the Utah Population Database. *Journal of the American Medical Informatics Association, 19*(e1), e54–e59. doi: [10.1136/amiajnl-2011-000335](https://doi.org/10.1136/amiajnl-2011-000335)
- Eijkemans, M. J., van Poppel, F., Habbema, D. F., Smith, K. R., Leridon, H., & te Velde, E. R. (2014). Too old to have children? Lessons from natural fertility populations. *Human Reproduction, 29*(6), 1304–1312. doi: [10.1093/humrep/deu056](https://doi.org/10.1093/humrep/deu056)
- Ellsworth, S. G. (1963). Utah's struggle for statehood. *Utah Historical Quarterly, 31*(1), 60–69.
- Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., . . . Marks, J. S. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study. *American Journal of Preventive Medicine, 14*(4), 245–258. doi: [10.1016/s0749-3797\(98\)00017-8](https://doi.org/10.1016/s0749-3797(98)00017-8)
- Gage, S. H., Munafo, M. R., & Davey Smith, G. (2016). Causal inference in Developmental Origins of Health and Disease (DOHaD) research. *Annual Review of Psychology, 67*, 567–585. doi: [10.1146/annurev-psych-122414-033352](https://doi.org/10.1146/annurev-psych-122414-033352)
- Gagnon, A., Smith, K. R., Tremblay, M., Vézina, H., Paré, P.-P., & Desjardins, B. (2009). Is there a trade-off between fertility and longevity? A comparative study of women from three large historical databases accounting for mortality selection. *American Journal of Human Biology, 21*(4), 533–540. doi: [10.1002/ajhb.20893](https://doi.org/10.1002/ajhb.20893)
- Garibotti, G., Smith, K. R., Kerber, R. A., & Boucher, K. M. (2006). Longevity and correlated frailty in multigenerational families. *The Journals of Gerontology: Series A, 61*(12), 1253–1261. doi: [10.1093/gerona/61.12.1253](https://doi.org/10.1093/gerona/61.12.1253)
- Greene, D., Tschanz, J. T., Smith, K. R., Østbye, T., Corcoran, C., Welsh-Bohmer, K. A., & Norton, M. C. (2014). Impact of offspring death on cognitive health in late life: The Cache County study. *The American Journal of Geriatric Psychiatry, 22*(11), 1307–1315. doi: [10.1016/j.jagp.2013.05.002](https://doi.org/10.1016/j.jagp.2013.05.002)
- Greeson, J. K., Briggs, E. C., Layne, C. M., Belcher, H. M. E., Ostrowski, S. A., Kim, S., . . . Fairbank, J. A. (2014). Traumatic childhood experiences in the 21st century: Broadening and building on the ACE studies with data from the National Child Traumatic Stress Network. *Journal of Interpersonal Violence, 29*(3), 536–556. doi: [10.1177/0886260513505217](https://doi.org/10.1177/0886260513505217)
- Hammad, I. A., Meeks, H., Fraser, A., Theilen, L. H., Esplin, M. S., Smith, K. R., & Varner, M. W. (2020). Risks of cause-specific mortality in offspring of pregnancies complicated by hypertensive disease of pregnancy. *American Journal of Obstetrics and Gynecology, 222*(1), 75.e1–75.e9. doi: [10.1016/j.ajog.2019.07.024](https://doi.org/10.1016/j.ajog.2019.07.024)

- Hanson, H. A., Leiser, C. L., Madsen, M. J., Gardner, J., Knight, S., Cessna, M., . . . Camp, N. J. (2020). Family study designs informed by tumor heterogeneity and multi-cancer pleiotropies: The power of the Utah Population Database. *Cancer Epidemiology, Biomarkers & Prevention*, 29(4), 807–815. doi: [10.1158/1055-9965.EPI-19-0912](https://doi.org/10.1158/1055-9965.EPI-19-0912)
- Hanson, M. A., Poston, L., & Gluckman, P. D. (2019). DOHaD: The challenge of translating the science to policy. *Journal of Developmental Origins of Health and Disease*, 10(3), 263–267. doi: [10.1017/S2040174419000205](https://doi.org/10.1017/S2040174419000205)
- Harrell, C. J., Smith, K. R., & Mineau, G. P. (2008). Are girls good and boys bad for parental longevity? The effects of sex composition of offspring on parental mortality past age 50. *Human Nature*, 19, 56–69. doi: [10.1007/s12110-008-9028-2](https://doi.org/10.1007/s12110-008-9028-2)
- Hawkes, K., & Smith, K. R. (2009). Brief communication: Evaluating grandmother effects. *American Journal of Physical Anthropology*, 140(1), 173–176. doi: [10.1002/ajpa.21061](https://doi.org/10.1002/ajpa.21061)
- Hawkes, K., & Smith, K. R. (2010). Do women stop early? Similarities in fertility decline in humans and chimpanzees. *Annals of the New York Academy of Sciences*, 1204(1), 43–53. doi: [10.1111/j.1749-6632.2010.05527.x](https://doi.org/10.1111/j.1749-6632.2010.05527.x)
- Hollingshaus, M. S., & Smith, K. R. (2015). Life and death in the family: Early parental death, parental remarriage, and offspring suicide risk in adulthood. *Social Science & Medicine*, 131, 181–189. doi: [10.1016/j.socscimed.2015.02.008](https://doi.org/10.1016/j.socscimed.2015.02.008)
- Horne, B. D., Camp, N. J., Muhlestein, J. B., & Cannon-Albright, L. A. (2006). Identification of excess clustering of coronary heart diseases among extended pedigrees in a genealogical population database. *American heart journal*, 152(2), 305–311. doi: [10.1016/j.ahj.2005.12.028](https://doi.org/10.1016/j.ahj.2005.12.028)
- Hsueh, Y., & Anderton, D. L. (1990). Temporal dimensions of the fertility transition: An age-period-cohort analysis of frontier fertility. *Sociological Perspectives*, 33(4), 447–464. doi: [10.2307/1389166](https://doi.org/10.2307/1389166)
- Hurdle, J. F., Smith, K. R., & Mineau, G. P. (2013). Mining electronic health records: An additional perspective. *Nature Reviews Genetics*, 14, 75–75. doi: [10.1038/nrg3208-c1](https://doi.org/10.1038/nrg3208-c1)
- Jennings, J. A., Sullivan, A. R., & Hacker, J. D. (2012). Intergenerational transmission of reproductive behavior during the demographic transition. *The Journal of Interdisciplinary History*, 42(4), 543–569. doi: [10.1162/jinh\\_a\\_00304](https://doi.org/10.1162/jinh_a_00304)
- Jones, J. H., & Bliege Bird, R. (2013). The marginal valuation of fertility. *Evolution and Human Behavior*, 35(1). doi: [10.1016/j.evolhumbehav.2013.10.002](https://doi.org/10.1016/j.evolhumbehav.2013.10.002)
- Jorde, L. B. (2001). Consanguinity and prereproductive mortality in the Utah Mormon population. *Human Heredity*, 52(2), 61–65. doi: [10.1159/000053356](https://doi.org/10.1159/000053356)
- Kauwe, J. S. K., Ridge, P. G., Foster, N. L., & Cannon-Albright, L. A. (2013). Strong evidence for a genetic contribution to late-onset Alzheimer's disease mortality: A population-based study. *PLoS ONE*, 8(10), e77087. doi: [10.1371/journal.pone.0077087](https://doi.org/10.1371/journal.pone.0077087)
- Kerber, R. A. (1995). Method for calculating risk associated with family history of a disease. *Genetic Epidemiology*, 12(3), 291–301. doi: [10.1002/gepi.1370120306](https://doi.org/10.1002/gepi.1370120306)
- Kerber, R. A., O'Brien, E., Smith, K. R., & Cawthon, R. M. (2001). Familial excess longevity in Utah genealogies. *The Journals of Gerontology: Series A*, 56(3), B130–B139. doi: [10.1093/gerona/56.3.b130](https://doi.org/10.1093/gerona/56.3.b130)
- Kerber, R., O'Brien, E., Smith, K. R., & Mineau, G. P. (2008). Familial aggregation of elderly cause-specific mortality: Analysis of extended pedigrees in Utah, 1904–2002. In T. Bengtsson & G. P. Mineau (Eds.), *International Studies in Population: Volume 7. Kinship and demographic behavior in the past* (pp. 243–258). Dordrecht: Springer. Available from [10.1007/978-1-4020-6733-4\\_11](https://doi.org/10.1007/978-1-4020-6733-4_11)
- Leiser, C. L., Taddie, M., Hemmert, R., Richards Steed, R., VanDerslice, J. A., Henry, K., . . . Hanson, H. A. (2020). Spatial clusters of cancer incidence: Analyzing 1940 census data linked to 1966–2017 cancer records. *Cancer Causes & Control*, 31, 609–615. doi: [10.1007/s10552-020-01302-3](https://doi.org/10.1007/s10552-020-01302-3)
- Lindahll-Jacobsen, R., Hanson, H. A., Oksuzyan, A., Mineau, G. P., Christensen, K., & Smith, K. R. (2013). The male-female health-survival paradox and sex differences in cohort life expectancy in Utah, Denmark, and Sweden 1850–1910. *Annals of Epidemiology*, 23(4), 161–166. doi: [10.1016/j.annepidem.2013.02.001](https://doi.org/10.1016/j.annepidem.2013.02.001)
- Lyman, E. L. (1998). Mormon leaders in politics: The transition to statehood in 1896. *Journal of Mormon History*, 24(2), 30–54. Available from <https://digitalcommons.usu.edu/mormonhistory/vol24/iss2/1>
- Lynch, K. A., Mineau, G. P., & Anderton, D. L. (1985). Estimates of infant mortality on the western frontier: The use of genealogical data. *Historical Methods*, 18(4), 155–164. doi: [10.1080/01615440.1985.10594160](https://doi.org/10.1080/01615440.1985.10594160)
- Maloney, T. N., Hanson, H., & Smith, K. R. (2014). Occupation and fertility on the frontier: Evidence from the state of Utah. *Demographic Research*, 30, 853–886. doi: [10.4054/DemRes.2014.30.29](https://doi.org/10.4054/DemRes.2014.30.29)
- McLellan, T., Jorde, L. B., & Skolnick, M. H. (1984). Genetic distances between the Utah Mormons and related populations. *American Journal of Human Genetics*, 36(4), 836–857.

- Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., . . . Skolnick, M. H. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*, 266(5182), 66–71. doi: [10.1126/science.7545954](https://doi.org/10.1126/science.7545954)
- Mineau, G. P., Bean, L. L., & Skolnick, M. (1979). Mormon demographic history II: The family life cycle and natural fertility. *Population Studies*, 33(3), 429–446. doi: [10.2307/2173890](https://doi.org/10.2307/2173890)
- Mineau, G. P., Smith, K. R., & Bean, L. L. (2002). Historical trends of survival among widows and widowers. *Social Science & Medicine*, 54(2), 245–254. doi: [10.1016/S0277-9536\(01\)00024-7](https://doi.org/10.1016/S0277-9536(01)00024-7)
- Mineau, G., Smith, K., & Bean, L. (2004). Adult mortality risks and religious affiliation: The role of social milieu in biodemographic studies. *Annales de Démographie Historique*, 108(2), 85–104. doi: [10.3917/adh.108.0085](https://doi.org/10.3917/adh.108.0085)
- Mineau, G. P., & Trussell, J. (1982). A specification of marital fertility by parents' age, age at marriage and marital duration. *Demography*, 19(3), 335–349. doi: [10.2307/2060975](https://doi.org/10.2307/2060975)
- Moorad, J. A. (2013). Multi-level sexual selection: Individual and family-level selection for mating success in a historical human population. *Evolution*, 67(6), 1635–1648. doi: [10.1111/evo.12050](https://doi.org/10.1111/evo.12050)
- Moorad, J. A., Promislow, D. E. L., Smith, K. R., & Wade, M. J. (2011). Mating system change reduces the strength of sexual selection in an American frontier population of the 19th century. *Evolution and Human Behavior*, 32(2), 147–155. doi: [10.1016/j.evolhumbehav.2010.10.004](https://doi.org/10.1016/j.evolhumbehav.2010.10.004)
- Moorad, J. A., & Wade, M. J. (2013). Selection gradients, the opportunity for selection, and the coefficient of determination. *The American Naturalist*, 181(3), 291–300. doi: [10.1086/669158](https://doi.org/10.1086/669158)
- Nam, C. B., & Boyd, M. (2004). Occupational status in 2000: Over a century of census-based measurement. *Population Research and Policy Review*, 23, 327–358. Available from [10.1023/B:POPU.0000040045.51228.34](https://doi.org/10.1023/B:POPU.0000040045.51228.34)
- Nam, C. B., & Powers, M. G. (1983). *The socioeconomic approach to status measurement*. Houston, TX: Cap and Gown Press.
- Neklason, D. W., Stevens, J., Boucher, K. M., Kerber, R. A., Matsunami, N., Barlow, J., . . . Burt, R. W. (2008). American founder mutation for attenuated familial adenomatous polyposis. *Clinical Gastroenterology Hepatology*, 6(1), 46–52. doi: [10.1016/j.cgh.2007.09.017](https://doi.org/10.1016/j.cgh.2007.09.017)
- Norton, M. C., Fauth, E., Clark, C. J., Hatch, D., Greene, D., Pfister, R., . . . Smith, K. R. (2016). Family member deaths across adulthood predict Alzheimer's disease risk: The Cache County Study. *International Journal of Geriatric Psychiatry*, 31(3), 256–263. doi: [10.1002/gps.4319](https://doi.org/10.1002/gps.4319)
- Norton, M. C., Hatch, D. J., Munger, R. G., & Smith, K. R. (2017). Family member deaths in childhood predict systemic inflammation in late life. *Biodemography and Social Biology*, 63(2), 104–115. doi: [10.1080/19485565.2017.1281099](https://doi.org/10.1080/19485565.2017.1281099)
- Norton, M. C., Østbye, T., Smith, K. R., Munger, R. G., & Tschanz, J. T. (2009). Early parental death and late-life dementia risk: Findings from the Cache County Study. *Age and Ageing*, 38(3), 340–343. doi: [10.1093/ageing/afp023](https://doi.org/10.1093/ageing/afp023)
- Norton, M. C., Smith, K. R., Østbye, T., Tschanz, J. T., Corcoran, C., Schwartz, S., . . . Welsh-Bohmer, K. A. (2010). Increased risk of dementia when spouse has dementia? The Cache County Study. *Journal of the American Geriatrics Society*, 58(5), 895–900. doi: [10.1111/j.1532-5415.2010.02806.x](https://doi.org/10.1111/j.1532-5415.2010.02806.x)
- Norton, M. C., Smith, K. R., Østbye, T., Tschanz, J. T., Schwartz, S., Corcoran, C., . . . Welsh-Bohmer, K. A. (2011). Early parental death and remarriage of widowed parents as risk factors for Alzheimer disease: The Cache County study. *The American Journal of Geriatric Psychiatry*, 19(9), 814–824. doi: [10.1097/JGP.0b013e3182011b38](https://doi.org/10.1097/JGP.0b013e3182011b38)
- O'Brien, E., Kerber, R., Smith, K., Mineau, G., Boucher, K., & Reed, D. L. (2007). Familial mortality in the Utah Population Database: Characterizing a human aging phenotype. *The Journals of Gerontology: Series A*, 62(8), 803–812. doi: [10.1093/gerona/62.8.803](https://doi.org/10.1093/gerona/62.8.803)
- O'Brien, E., Rogers, A. R., Beesley, J., & Jorde, L. B. (1994). Genetic structure of the Utah Mormons: Comparison of results based on RFLPs, blood groups, migration matrices, isonymy, and pedigrees. *Human Biology*, 66(5), 743–759.
- Penkler, M., Hanson, M., Biesma, R., & Muller, R. (2019). DOHaD in science and society: Emergent opportunities and novel responsibilities. *Journal of Developmental Origins of Health and Disease*, 10(3), 268–273. doi: [10.1017/S2040174418000892](https://doi.org/10.1017/S2040174418000892)
- Penn, D. J., & Smith, K. R. (2007). Differential fitness costs of reproduction between the sexes. *Proceedings of the National Academy of Sciences of the United States of America*, 104(2), 553–558. doi: [10.1073/pnas.0609301103](https://doi.org/10.1073/pnas.0609301103)
- Rich, E. C., Burke, W., Heaton, C. J., Haga, S., Pinsky, L., Short, M. P., & Acheson, L. (2004). Reconsidering the family history in primary care. *Journal of the General Internal Medicine*, 19(3), 273–280. doi: [10.1111/j.1525-1497.2004.30401.x](https://doi.org/10.1111/j.1525-1497.2004.30401.x)

- Robson, S. L., & Smith, K. R. (2011). Twinning in humans: Maternal heterogeneity in reproduction and survival. *Proceedings of the Royal Society B*, 278(1725), 3755–3761. doi: [10.1098/rspb.2011.0573](https://doi.org/10.1098/rspb.2011.0573)
- Robson, S. L., & Smith, K. R. (2012). Parity progression ratios confirm higher lifetime fertility in women who bear twins. *Proceedings of the Royal Society B*, 279(1738). doi: [10.1098/rspb.2012.0436](https://doi.org/10.1098/rspb.2012.0436)
- Sanz-de-Galdeano, A., Terskaya, A., & Upegui, A. (2020). Association of a genetic risk score with BMI along the life-cycle: Evidence from several US cohorts. *PLoS ONE*, 15(9), e0239067. doi: [10.1371/journal.pone.0239067](https://doi.org/10.1371/journal.pone.0239067)
- Schacht, R., Meeks, H., Fraser, A., & Smith, K. R. (2021). Was Cinderella just a fairy tale? Survival differences between stepchildren and their half-siblings. *Philosophical Transactions of the Royal Society B*, 376(1827), 20200032. Available from [10.1098/rstb.2020.0032](https://doi.org/10.1098/rstb.2020.0032)
- Schacht, R., & Smith, K. R. (2017). Causes and consequences of adult sex ratio imbalance in a historical U.S. population. *Philosophical Transactions of the Royal Society B*, 372(1729). doi: [10.1098/rstb.2016.0314](https://doi.org/10.1098/rstb.2016.0314)
- Schacht, R., Tharp, D., & Smith, K. R. (2019). Sex ratios at birth vary with environmental harshness but not maternal condition. *Scientific Reports*, 9, 9066. doi: [10.1038/s41598-019-45316-7](https://doi.org/10.1038/s41598-019-45316-7)
- Scheuner, M. T., Wang, S.-J., Raffel, L. J., Larabell, S. K., & Rotter, J. I. (1997). Family history: A comprehensive genetic risk assessment method for the chronic conditions of adulthood. *American Journal of Medical Genetics*, 71(3), 315–324. doi: [10.1002/\(sici\)1096-8628\(19970822\)71:3<315::aid-ajmg12>3.0.co;2-n](https://doi.org/10.1002/(sici)1096-8628(19970822)71:3<315::aid-ajmg12>3.0.co;2-n)
- Sellers, T. A., Caporaso, N., Lapidus, S., Petersen, G. M., & Trent, J. (2006). Opportunities and barriers in the age of team science: Strategies for success. *Cancer Causes & Control*, 17(3), 229–237. doi: [10.1007/s10552-005-0546-5](https://doi.org/10.1007/s10552-005-0546-5)
- Shah, R., Pico, A. R., & Freedman, J. E. (2016). Translational epidemiology: Entering a brave new world of team science. *Circulation Research*, 119(10), 1060–1062. doi: [10.1161/CIRCRESAHA.116.309881](https://doi.org/10.1161/CIRCRESAHA.116.309881)
- Skolnick, M., Bean, L., May, D., Arbon, V., De Nevers, K., & Cartwright, P. (1978). Mormon demographic history I. Nuptiality and fertility of once-married couples. *Population Studies*, 32(1), 5–19. doi: [10.1080/00324728.1978.10412788](https://doi.org/10.1080/00324728.1978.10412788)
- Smith, K. R., Brown, B. B., Yamada, I., Kowaleski-Jones, L., Zick, C. D., & Fan, J. X. (2008). Walkability and body mass index density, design, and new diversity measures. *The American Journal of Preventive Medicine*, 35(3), 237–244. doi: [10.1016/j.amepre.2008.05.028](https://doi.org/10.1016/j.amepre.2008.05.028)
- Smith, K. R., Gagnon, A., Cawthon, R. M., Mineau, G. P., Mazan, R., & Desjardins, B. (2009). Familial aggregation of survival and late female reproduction. *The Journals of Gerontology: Series A*, 64A(7), 740–744. doi: [10.1093/gerona/glp055](https://doi.org/10.1093/gerona/glp055)
- Smith, K. R., Hanson, H., & Mineau, G. (2016). A six-pack of big ideas for historical demography. In K. Matthijs, S. Hin, J. Kok, & H. Matsuo (Eds.), *The future of historical demography* (pp. 211–214). Leuven/Den Haag: Acco.
- Smith, K. R., Hanson, H. A., Mineau, G. P., & Buys, S. S. (2012). Effects of BRCA1 and BRCA2 mutations on female fertility. *Proceedings of the Royal Society B*, 279(1732), 1389–1395. doi: [10.1098/rspb.2011.1697](https://doi.org/10.1098/rspb.2011.1697)
- Smith, K. R., Hanson, H. A., Norton, M. C., Hollingshaus, M. S., & Mineau, G. P. (2014). Survival of offspring who experience early parental death: Early life conditions and later-life mortality. *Social Science & Medicine*, 119, 180–190. doi: [10.1016/j.socscimed.2013.11.054](https://doi.org/10.1016/j.socscimed.2013.11.054)
- Smith, K. R., Mineau, G. P., & Bean, L. L. (2002). Fertility and post-reproductive longevity. *Social Biology*, 49(3–4), 185–205.
- Smith, K. R., Mineau, G. P., Garibotti, G., & Kerber, R. (2009). Effects of childhood and middle-adulthood family conditions on later-life mortality: Evidence from the Utah Population Database, 1850–2002. *Social Science & Medicine*, 68(9), 1649–1658. doi: [10.1016/j.socscimed.2009.02.010](https://doi.org/10.1016/j.socscimed.2009.02.010)
- Smith, K. R., Zick, C. D., Kowaleski-Jones, L., Brown, B. B., Fan, J. X., & Yamada, I. (2011). Effects of neighborhood walkability on healthy weight: Assessing selection and causal influences. *Social Science Research*, 40(5), 1445–1455. doi: [10.1016/j.ssresearch.2011.04.009](https://doi.org/10.1016/j.ssresearch.2011.04.009)
- Song, X., & Campbell, C. D. (2017). Genealogical microdata and their significance for social science. *Annual Review of Sociology*, 43, 75–99. doi: [10.1146/annurev-soc-073014-112157](https://doi.org/10.1146/annurev-soc-073014-112157)
- Stokols, D., Misra, S., Moser, R. P., Hall, K. L., & Taylor, B. K. (2008). The ecology of team science: Understanding contextual influences on transdisciplinary collaboration. *American Journal of Preventive Medicine*, 35(2 Suppl), S96–115. doi: [10.1016/j.amepre.2008.05.003](https://doi.org/10.1016/j.amepre.2008.05.003)
- Stroup, A. M., Herget, K. A., Hanson, H. A., Reed, D. L., Butler, J. T., Henry, K. A., . . . Smith, K. R. (2017). Baby Boomers and birth certificates: Early-life socioeconomic status and cancer risk in adulthood. *Cancer Epidemiology, Biomarkers & Prevention*, 26(1), 75–84. doi: [10.1158/1055-9965.EPI-16-0371](https://doi.org/10.1158/1055-9965.EPI-16-0371)
- Temby, O. F., & Smith, K. R. (2014). The association between adult mortality risk and family history of longevity: The moderating effects of socioeconomic status. *Journal of Biosocial Science*, 46(6), 703–716. doi: [10.1017/s0021932013000515](https://doi.org/10.1017/s0021932013000515)

- Theilen, L. H., Fraser, A., Hollingshaus, M. S., Schliep, K. C., Varner, M. W., Smith, K. R., & Esplin, M. S. (2016). All-cause and cause-specific mortality after hypertensive disease of pregnancy. *Obstetrics & Gynecology*, 128(2), 238–244. doi: [10.1097/AOG.0000000000001534](https://doi.org/10.1097/AOG.0000000000001534)
- Theilen, L. H., Meeks, H., Fraser, A., Esplin, M. S., Smith, K. R., & Varner, M. W. (2018). Long-term mortality risk and life expectancy following recurrent hypertensive disease of pregnancy. *American Journal of Obstetrics & Gynecology*, 219(1), P107.E101-107.E6. doi: [10.1016/j.ajog.2018.04.002](https://doi.org/10.1016/j.ajog.2018.04.002)
- Trivers, R. L., & Willard, D. E. (1973). Natural selection of parental ability to vary the sex ratio of offspring. *Science*, 179(4068), 90–92. doi: [10.1126/science.179.4068.90](https://doi.org/10.1126/science.179.4068.90)
- Tschanz, J. T., Pfister, R., Wanzek, J., Corcoran, C., Smith, K., Tschanz, B. T., . . . Norton, M. C. (2013). Stressful life events and cognitive decline in late life: Moderation by education and age. The Cache County Study. *International Journal of Geriatric Psychiatry*, 28(8), 821–830. doi: [10.1002/gps.3888](https://doi.org/10.1002/gps.3888)
- van den Berg, N., Beekman, M., Smith, K. R., Janssens, A., & Slagboom, P. E. (2017). Historical demography and longevity genetics: Back to the future. *Ageing Research Reviews*, 38, 28–39. doi: [10.1016/j.arr.2017.06.005](https://doi.org/10.1016/j.arr.2017.06.005)
- van den Berg, N., Rodríguez-Girondo, M., van Dijk, I. K., Mourits, R. J., Mandemakers, K., Janssens, A. A. P. O., . . . Slagboom, P. E. (2019). Longevity defined as top 10% survivors and beyond is transmitted as a quantitative genetic trait. *Nature Communications*, 10(1), 35. doi: [10.1038/s41467-018-07925-0](https://doi.org/10.1038/s41467-018-07925-0)
- van Dijk, I. K., Janssens, A., & Smith, K. R. (2019). The long harm of childhood: Childhood exposure to mortality and subsequent risk of adult mortality in Utah and the Netherlands. *European Journal of Population*, 35(5), 851–871. doi: [10.1007/s10680-018-9505-1](https://doi.org/10.1007/s10680-018-9505-1)
- Zick, C. D., Smith, K. R., Fan, J. X., Brown, B. B., Yamada, I., & Kowaleski-Jones, L. (2009). Running to the store? The relationship between neighborhood environments and the risk of obesity. *Social Science & Medicine*, 69(10), 1493–1500. doi: [10.1016/j.socscimed.2009.08.032](https://doi.org/10.1016/j.socscimed.2009.08.032)
- Zimmer, Z., Hanson, H., & Smith, K. (2016a). Childhood socioeconomic status, adult socioeconomic status, and old-age health trajectories: Connecting early, middle, and late life. *Demographic Research*, 34, 285–320. doi: [10.4054/DemRes.2016.34.10](https://doi.org/10.4054/DemRes.2016.34.10)
- Zimmer, Z., Hanson, H. A., & Smith, K. R. (2016b). Offspring socioeconomic status and parent mortality within a historical population. *Demography*, 53(5), 1583–1603. doi: [10.1007/s13524-016-0502-x](https://doi.org/10.1007/s13524-016-0502-x)

HISTORICAL LIFE COURSE STUDIES  
VOLUME 11 (2021), published 21-09-2021

# The Long Road to Health and Prosperity, Southern Sweden, 1765–2015

Research Contributions From the Scanian Economic-  
Demographic Database (SEDD)

Tommy Bengtsson  
Lund University

Martin Dribe  
Lund University

## ABSTRACT

The Scanian Economic-Demographic Database (SEDD) at the Centre for Economic Demography (CED), Lund University was built to answer questions derived from previous research using macro data from 1749 onwards. It includes longitudinal micro data for a regional sample of rural, semi-urban, and urban parishes in southern Sweden from 1646 to 1968 for approximately 175,000 individuals. In addition to the data on births, deaths, marriages, and occupations, it includes data on migration, household size, landholdings, taxation, and heights from the 1800s onwards and on income from 1865 onwards. After being linked from 1968 to 2015 to a range of national registers with detailed demographic and socioeconomic information, it includes 825,000 individuals. The richness and wide range of micro data have allowed researchers to follow individuals throughout their lives and across generations, covering extensive periods, and to make comparisons with results from macro data. This research has partly confirmed the established view on long-term changes in living standards and demographics in Sweden but has also brought into question some previously held truths.

**Keywords:** Sweden, Demographic transition, Family reconstitution, Population registers, Longitudinal data, Life courses, Health, Living standards

DOI article: <https://doi.org/10.51964/hlcs10941>

© 2021, Bengtsson, Dribe

This open-access work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/), which permits use, reproduction & distribution in any medium for non-commercial purposes, provided the original author(s) and source are given credit. See <http://creativecommons.org/licenses/>.

## 1 INTRODUCTION

The Scanian Economic-Demographic Database (SEDD) was initially compiled to answer the questions that stemmed from previous research using tabular data from 1749, the year when population censuses started in Sweden, and onwards. These analyses covered either the entire country, counties, or smaller units. Since the population tables cannot be used to answer questions about the roles of income and wealth, social position, and household composition for demographic outcomes, SEDD was used instead. Today SEDD includes data for a regional sample of rural, semi-urban, and urban parishes in southern Sweden from 1646 to 1968 for approximately 175,000 individuals. From 1968 to 2015, SEDD has been linked to a range of national registers with detailed demographic and socioeconomic information on all individuals who have ever lived in the area and their partners, children and grandchildren. The linked micro database includes 825,000 individuals. As SEDD has continuously expanded over the past 30 years in terms of the geographical area and time period covered, not all studies are based on exactly the same sample. A detailed description of the sample, the sources and the data structure is presented in Dribe and Quaranta (2020). The income data is presented in Helgertz, Bengtsson and Dribe (2020). The fact that Sweden started much earlier than other countries making population censuses allows us to make comparisons with national development during the preindustrial era.

Although SEDD is not a representative sample of Sweden, demographic development in the area is similar to that nationally. For example, trends in both life expectancy and fertility are similar in the study area and in Sweden as a whole (Bengtsson & Dribe, 1997; Lazuka, 2017). The mortality response to short-term economic stress is also similar to the national pattern both overall and by age and sex (Bengtsson, 2004b; Bengtsson & Dribe, 2005; Bengtsson & Ohlsson, 1994). The same is true for the fertility response to short-term economic stress (Bengtsson, 2000; Bengtsson & Dribe, 2006).

The changes in the population totals for Scania were also similar to those for the entire country and of several other European countries for which pre-census estimations have been made. In fact, the estimations of the changes in the population totals for Scania show similarities to England back to 1650 (Bengtsson & Oeppen, 1993). While population development was similar to that of other European countries, economic development differed considerably. Sweden had its industrial take-off with increasing real wages in the 1870s, about a century later than England (Bengtsson & Jörberg, 1975; Crafts & Mills, 2020; Wrigley, 2011).

Although real wages in Sweden showed no clear trend before the industrial take-off, they were not stable annually. On the contrary, real wages varied greatly by year. This variation in real wages, stemming from changes in food prices, had a strong short-term impact on the fertility and mortality rates and a slightly weaker impact on the nuptiality rates. The mortality impact decreased during the 19th century (Bengtsson & Ohlsson, 1978, 1984). In this respect, Swedish development was similar to that of England (Lee, 1981), with the difference that it took place later in Sweden, just as industrialisation did. However, the variation in real wages and food prices did not affect the mortality rate at all ages. Paradoxically, while the children and adults, both male and female, suffered heavily, the elderly people were less affected, and the infants were not affected at all (Bengtsson & Ohlsson, 1985). Thus, our research based on the aggregated national data confirmed the Malthusian view that in the 18th century, many people lived close to the subsistence level (Heckscher, 1949; Malthus, 1803; Wargentin, 1772/1976). More interestingly, the findings showed, contrary to what Malthus anticipated, that in the beginning of the 19th century, the mortality response to food prices became weaker despite rapid population growth. This phenomenon meant that at least some groups escaped hunger and premature death, to use Fogel's phrase (Fogel, 1996). A recent study using parish data for southern Sweden supports this conclusion (Dribe, Olsson, & Svensson, 2017).

According to the traditional view, the demographic transition in Sweden started in approximately 1800 with a decline in the crude death rate due to modernization and continued with a decline in the crude birth rate in the 1880s (e.g., Notestein, 1953). The problem is that there was not much modernisation in 1800, apart from improvements in agriculture. Real wages did not start to increase until the 1870s (see Bengtsson & Ohlsson, 1994). This means that there was a mismatch in the timing between the mortality rate decline and modernisation (e.g. industrialisation). In fact, the mortality rate started to decline well before 1800 due to a reduction in smallpox mortality rates among children, making the gap even wider (Bengtsson, 2001). This implies that the first stage of the mortality decline in Sweden, the reduction of the infant and child mortality rates, was part of an older pattern in which the level of mortality varied independently of modernisation, as in England (Wrigley & Schofield, 1981). Only after

1870, when the adult mortality rate also declined, there was a direct association between economic development and the mortality rate, and at approximately the same time, the fertility rate started to decline. For this reason, Bengtsson and Ohlsson (1994) argued that the demographic transition in Sweden did not start with the mortality decline in 1800, but in the 1870s, with modernisation, a continued decline occurred in the mortality rate at all ages, and the fertility rate began to decline.

Taken together, the results from aggregated studies on the mortality and fertility rates, and real wages, confirm the Malthusian view of the poor conditions in the 18th century, but not his predictions that conditions got even worse when population grew rapidly, and question the theory of the demographic transition. To understand the changes in the living conditions that occurred when Sweden transformed from an agricultural to an industrial society, it is necessary to go beyond aggregated statistics. As a first step, more knowledge is needed about how different social and economic groups were affected by this transformation. For this reason, Tommy Bengtsson and Rolf Ohlsson, in 1983, started a collaboration with the Regional Archives in Lund with the aim of developing a database linking births, deaths, and marriages from church books with information on occupations for a sample of nine parishes and a town in western Scania, which later became the SEDD. Bengtsson collaborated with statistician Göran Broström on how to integrate the analyses of aggregated data on real wages and food prices with the individual-level demographic data to connect to our previous research (see, e.g., Bengtsson, 1997a, 1999; Bengtsson & Broström, 2010).

With the start of the comparative *EurAsian Project on Population and Family History* in 1993, SEDD was expanded with household information from catechetical examination records as well as from income and other records for five of the original nine parishes and, recently, for the port town of Landskrona. The data date back to 1646 for two parishes, 1686 for three parishes, and, presently, 1905 for Landskrona (see Dribe & Quaranta, 2020). Using the linked data, the area population can be followed without gaps from about 1815 to 2015, with information on demographic, economic, and social conditions. Data on migration are also available from approximately 1815, which means that information on exposure is available for or the entire population from this year. Most importantly, the demographic development in this area — including the changes in total population, life expectancy, and family size — follows the same time trends as the entire country (Bengtsson & Dribe, 1997; Lazuka, 2017; Quaranta, 2013).

Below, we overview the findings from the research based on the SEDD. The richness and wide range of the data allow us to conduct research covering extensive periods and many topics, following individuals along their life courses and across generations. We start with an overview of the economic development and social structure of the area and continue with topics related to the demographic transition. We summarise the research on how conditions in early life influenced health and well-being later in life, the delayed effects of conditions in early life. We start each section by giving a brief overview of the context and previous research at the national and regional levels to compare with the findings using the SEDD microdata where they overlap.

## 2 ECONOMIC DEVELOPMENT AND SOCIAL STRUCTURE

During the 19th century, the transformation from an agricultural to an industrial economy — together with the demographic transition, urbanisation, and large-scale emigration — fundamentally changed the living conditions for the Swedish population. The public poor-law system lagged behind, as it did in other countries at the time (Bengtsson, 2004b; Skoglund, 1992). It provided help only to disabled and elderly people, and since the manorial lords assisted only their own employees, most people had to rely on their families if they could not provide for themselves (Dribe, Olsson, & Svensson, 2010; Lundh & Olsson, 2002, 2009). As a result, in bad years, theft and crime increased (Hellstenius, 1871), and hordes of people left their homes to find work elsewhere (Utterström, 1957). Some went to towns, while others went to nearby villages (Bengtsson, 1987, 1990; Bengtsson & Dribe, 1997).

The first step to develop the welfare state was taken when a pension system was introduced in 1913. Social welfare expanded in the 1930s, and in 1948 pensions covered the costs-of-living for the first time, but it was not until the 1960s and thereafter that the welfare systems became more extensive (Elmér, 1971). Meanwhile, school participation rose. The percentage of all 16-year-olds who had

completed at least nine years of schooling increased from approximately 4% in 1930 to approximately 26% in 1965 (Stanfors, 2007), and since the 1960s, university education has expanded rapidly.

Scania is known as the granary of Sweden. When Sweden acquired the province from Denmark in 1658, the noble class owned 54% of the land, the crown owned 29%, the church owned 11%, and the freeholders owned 8% (Weibull, 1923). As tenants bought land from the crown in the 18th century at favourable prices, their share of the land increased steeply. For example, in Hög and Kävlinge, two of the five rural parishes included in the SEDD, the freeholders owned 10% of the land in 1730, 50% in 1800, and 95% in 1870 (Svensson, 2013). This shift occurred as agricultural reforms were being made. In the five rural parishes, enclosure reforms took place between 1764 and 1849, except on one of the estates, where it did not happen until 1914 (Bengtsson & Dribe, 1997). Together with land reclamation, new crops, new technology, labour reorganisation, and access to foreign markets, production increased in both absolute terms and per capita (Olsson & Svensson, 2010). In parallel, the number of farm workers increased, and by the beginning of the 19th century, more than half of the population was landless (Bengtsson, 2004b; Bengtsson & Dribe, 1997).

As the accumulation of land and the division of farms occurred in parallel, the structure of the farmer class changed. The farmers became a heterogeneous group in terms of farm size. While some farmers hired labourers to work for them, others were themselves part-time labourers (Bengtsson & Dribe, 1997). Household size, likewise, varied greatly depending on land size (Dribe, 2000; Lundh, 1995).

When the railroad network expanded in the 1860s, one of the rural parishes, Kävlinge, became a hub and changed into a small industrial town characterised by the food and textile industries. The other four parishes — Hög, Halmstad, Sireköpinge and Kågeröd — remained largely rural. Landskrona, a port town involved in the grain trade, became an important place for shipbuilding and other industrial activities (Dribe & Svensson, 2019).

Preindustrial rural societies have often been characterised as stationary and immobile, both geographically and socially. Several studies using the SEDD have proven this view wrong. Instead, social mobility in the five parishes was high (Dribe & Svensson, 2008a). Downwards mobility, e.g., from the peasant class to the semi-landless and landless groups, was more frequent than upward mobility and increased over the 19th century (see also Lundh, 1999a). Social attainment was determined by both inheritance and individual agency. Social and spousal origins were of crucial importance for socioeconomic attainment in this largely rural context (Dribe & Lundh, 2009a). Both social homogamy and geographic endogamy play important roles in socioeconomic reproduction (Dribe & Lundh, 2010).

Before 1900, access to land was the main basis of someone's social position in the countryside. Families adopted strategies to secure the transmission of land from one generation to the next, despite a rather equal inheritance legislation under which all sons and daughters inherited equal shares from 1857 onwards, and where previously sons had inherited twice the amount that daughters did. The families in the rural areas between 1720 and 1840 often used formal retirement contracts to circumvent the inheritance legislation to pass on the landholding to one of the children (Dribe & Lundh, 2005a). Based on the post-mortem inventories, Dribe and Lundh (2005a) found that, while sons were more likely to take over, daughters (rather, sons-in-law) often took over as well, and not only in cases where no sons were available. This practice indicates the flexible strategies used in the intergenerational transmission of land, the most important productive resource in this context. Land transmission was crucial for securing reproduction, as it gave the new generation access to marriage and childbearing. Retirement contracts were also important among manorial tenants, who did not own their land. However, the conditions were usually not as beneficial as they were among freeholders, and the noble landowners sometimes intervened in the intergenerational transmission of the tenancies (Lundh & Olsson, 2002).

As the land market developed in a capitalist direction during the 19th century, these land transmission strategies changed even though retirement remained an important strategic aspect (Dribe & Lundh, 2005b). Earlier, land transmission was largely an intra-family affair, in which the value of the property was kept low to facilitate transmission to one chosen heir. In the first half of the 19th century, it was increasingly channelled through the market as the real value of the property became clear to all heirs.

When looking at the 20th century, land becomes insufficient to assess socioeconomic status and must be complemented by other aspects. Occupation-based social class offers a broader view of social stratification, and in a number of studies, such class schemes have been used to study both socioeconomic mobility and socioeconomic differentials in demographic outcomes such as marriage

and the fertility and mortality rates. Two different, but in many ways quite similar, class schemes have been employed in different studies: SOCPO (Van de Putte & Miles, 2005; Van de Putte & Svensson, 2010) and HISCLASS (van Leeuwen & Maas, 2011). They are both based on the occupations coded in HISCO (van Leeuwen, Maas, & Miles, 2002). The recent SEDD releases include the occupational notations coded in HISCO and coding schemes to derive HISCLASS (see Dribe & Quaranta, 2020).

The data for the five parishes for 1815–1968 have been used to analyse the long-term development of social mobility based on the SOCPO classification (Dribe, Helgertz, & Van de Putte, 2015). The analysis was related to the extensive scholarship in sociology, which had debated whether social mobility increased from preindustrial to industrial times and the extent to which social mobility differed across industrial societies depending on the level of development (e.g., Erikson & Goldthorpe, 1992; Featherman, Jones, & Hauser, 1975; Grusky & Hauser, 1984; Lipset & Zetterberg, 1956). Both absolute (total) and relative (societal openness) mobility increased over time, mainly as a result of the increasing upward mobility in the 20th century, as educational expansion promoted the entry of people from working-class origins to the expanding middle class of white-collar workers.

Expanding the analysis of social mobility to three generations in the five parishes between 1815 and 2011, based on HISCLASS, Dribe and Helgertz (2016) demonstrated a clear association between the grandfathers' and the grandsons' classes, net of the class of the fathers. They also analysed the intergenerational associations in occupational status using the HISCAM scale (Lambert, Zijdemans, van Leeuwen, Maas, & Prandy, 2013) and earnings. While the HISCAM associations were similar to those found in HISCLASS, they found no association between the earnings of grandfathers and grandsons net of the earnings of the father. The literature includes a debate about how to interpret these kinds of 'grandfather effects': whether they represent a real influence from grandfathers to grandsons or whether they result from measurement errors or random shocks to attainment in the father generation.

A recent study of higher education attainment used the data for the five parishes and Landskrona for 1939–2015 to study the role of the socioeconomic composition of the neighbourhood (Hedefalk & Dribe, 2020). For this study, the individual-level data were geocoded at the address level and used to create a flexible measure of the socioeconomic status of the neighbours. The results showed that the neighbourhood social class in childhood was associated with attaining a higher education regardless of the social origin and the school the children attended.

To summarise, research using the SEDD has demonstrated that there was considerable social mobility in preindustrial society, especially downwards from the farmer class to landless or semi-landless labourers. As the area industrialised, social mobility increased, and later in the industrialisation process, upward mobility increased, connected to the increasing importance of education and merit-based recruitment in the labour market.

### 3 MIGRATION

Preindustrial society has often been viewed as geographically immobile. Especially within modernisation theory, which became popular in the social sciences in the post-WWII period, rural society has often been pictured as stationary and the people who lived there as unwilling to move. Only with industrialisation and the connection of local markets to wider regional, national and later international markets did conditions develop for migration (see, e.g., Hochstadt, 1999). In a number of studies using the SEDD, this picture of an immobile preindustrial society has been refuted. Geographical mobility was high in preindustrial society. In fact, people moved almost as frequently in the 19th century as they do today (Dribe, 2003a).

In the SEDD area, during the first half of the 19th century, only 30% of the population over age 25 were born in the same parish as the one in which they resided (Dribe, 2000). On the other hand, migration was local, and almost 80% lived within 15 kilometres from their place of birth. Overall, migration occurred primarily within a 15-kilometre range: 87% for children leaving home, 86% for family migration and 78% for servants (Dribe, 2003b). Most migrants were young adults, and servants especially were mobile, often moving every year. The high mobility of servants was related not only to work organisation and occupational mobility but also to marriage, dissatisfaction with employers, or economic fluctuations affecting the demand for labour (Dribe & Lundh, 2005c; Lundh, 1999b).

Additionally, families moved quite frequently (Dribe, 2003b), especially those belonging to the group of landless labourers (Dribe & Svensson, 2008b).

Areas with rapid agricultural development in the first half of the 19th century and, later, areas with industrialisation offered higher wages to attract labour, which led to an increase in longer range, rural-to-urban migration (Bengtsson, 1990). As people migrated to these areas to meet the demand for labour, the wage differentials eventually levelled off and migration declined.

Migration could also be a way to mitigate the impact of short-term economic stress caused by local harvest failures or external changes in food prices (Allen, Bengtsson, & Dribe, 2005; Bengtsson, 2004a). Although one could expect people to leave when conditions were bad, research based on SEDD has shown that migration was a quite inefficient relief in times of economic stress, especially for those most in need (Bengtsson, 1987; Dribe, 2000, 2003c). The landed groups economised on labour, adjusting the timing of their sons' leaving home to economic conditions (Dribe 2000, 2004a). In years with low prices, when they obtained lower revenues selling their products, they kept their children at home as a substitute for servants, which minimised their labour costs. This fact means that the migration of children from the parental home was an important part of the strategic actions of the peasant families in balancing the demand and supply of the household labour (Dribe, 2000; Dribe & Lundh, 2005c).

Regardless of food prices, a large majority of the children of landless and semi-landless labourers left home around age 15, when their family, or employer, had to start paying taxes for them. Even though these families suffered from high food prices, they were unable to cope with their difficulties by letting their children move away from the household. Instead, the landed group adjusted their family sizes and compositions in years of economic stress (Dribe, 2000, 2003c, 2004a). Patterns of leaving home were also strongly affected by demographic stress in the household, stemming from the death of one of the parents. Such events were largely disruptive and had similar effects on sons and daughters, with the exception of older daughters with younger siblings, who tended to remain at home longer after the deaths of their mothers (Dribe, 2000).

A number of studies using the SEDD have shown that preindustrial society was not as immobile as has previously been argued. In fact, people moved almost as frequently in the 19th century as they do today. Migration, temporary and permanent, has been seen as a way of reducing the stress caused by variations in harvests and food prices. The landed families were better able than the landless families to take advantage of this instrument. The farmers could economise with the labour of their children, since in bad years, the children could stay at home to work on the farm. Thus, migration was not the universal solution to short-term economic stress, as has previously been thought.

## 4 MARRIAGE

The household was of the utmost importance in preindustrial times. Living alone was not an option. The division of labour within the household secured a couple's well-being in different ways. Following Peter Laslett's research on the predominance of nuclear-family households in preindustrial England, in the 1970s and 1980s, interest surged in historical family patterns (e.g., Berkner, 1972; Hammel & Laslett, 1974; Laslett, 1965/1994; Laslett & Wall, 1972). Studies using early SEDD data have shown that household size and composition differed by social group (Dribe, 2000; Lundh, 1995). A trend was found towards smaller households from the mid-18th to the mid-19th century, partly because of proletarianisation, through which the social groups with the smallest households increased their proportion of all households. The size and structure of the households, especially the peasant households, also changed over the family life cycle, and servants were used as substitutes in times when the supply of family labour was low (Dribe, 2000; Lundh, 1995).

In Sweden, as in the rest of Western Europe, a new independent household was formed when a couple married (Hajnal, 1983; see also Lundh & Kurosu, 2014). In this way, marriage was an important life-course transition between the leaving of the parental home and the birth of the first child, and the incidence of bridal pregnancy was high in many contexts, including Scania (Dribe, Manfredini, & Oris, 2014). In the life-course trajectory, socioeconomic status differed considerably, both with regard to the age at leaving home and the age at marriage.

Sweden had a Western European marriage pattern, characterised by late marriages and high population proportions never marrying (Hajnal, 1965). In 1900, 14% of the men and 19% of the women aged 45 to 49 had never been married (Dribe & Lundh, 2014). Regional variations were considerable in the marriage patterns in Sweden (Lundh, 1993, 1999c, 2013). Marriage patterns between the socioeconomic groups were distinctly different for the men but not for the women. At age 45, 6% of the landed men and 25% of the landless men were unmarried in the 19th century (Dribe & Lundh, 2014). Marriage occurred late, for both the men and the women. The landed males married on average at age 27, and the landless men did so at age 28. The women married on average at age 25 regardless of socioeconomic status.

Important were both when to marry and to whom. Choosing the right spouse was crucial, especially for the landholding population, and marriage was largely a strategic alliance to secure the social reproduction of the landholding class. Scania, as were most other preindustrial rural areas of the time, was characterised by a strong social homogamy, especially among the landowners. They were much more likely to marry a spouse from the same social origin than were those of landless or semi-landless origins. This strong pattern of social homogamy did not change much over the 19th century (Dribe & Lundh, 2005d). Age homogamy and geographic endogamy were also strong but not as strong as social homogamy, and geographic endogamy became especially less prevalent over the 19th century (Dribe & Lundh, 2009b).

Moreover, the timing of marriage during the year was important in an agricultural society in which the labour demand varied strongly by season. Getting married during the peak harvest season was impossible, as most people would have been fully occupied with farm work during that time. By looking at the changes in marriage seasonality in Scania between 1685 and 1895, Dribe and Van de Putte (2012) traced an increase in the work intensity over the year, consistent with the idea of an 'industrious revolution' (de Vries, 1994). The seasonality of marriage changed dramatically over time, as marriages became increasingly concentrated in the remaining slack season in December, especially around Christmas.

The death of a spouse brought profound changes to the living conditions of the surviving party. This phenomenon caused remarriages to be an integral part of the marriage system (Lundh, 2002, 2007; Kurosu, Lundh, & Breschi, 2014). The consequences of the decline in support could be disastrous, though various kinds of compensation were available, such as inheritance, gifts and charity (economic support), hired domestic servants (service support), and children and other relatives or neighbours and friends (social and emotional support).

Access to such compensatory measures in a peasant society depended on factors such as gender, household structure, and socioeconomic status. Widows, especially if they were landless, were more dependent on economic assistance to offset the loss of a deceased husband's income. Widowers were more in need of extra service support to replace the domestic work of a dead wife (Nystedt, 2002). The presence of children in the household was a potential source of all types of support (Dribe, Lundh, & Nystedt, 2007). The men were more likely to remarry than the women were, and the young people were more likely to do so than the old people were. Remarriages also became less common over time, and, while the likelihood of remarriage among men exhibited no socioeconomic difference, such a difference was found among the women. A peasant widow was approximately 50% more likely to remarry than a non-peasant widow was.

Changes in the food prices seem to have had no effect on the propensity to remarry among the farmers, although they had a clear *negative* effect on the non-farmers in the 19th century (Lundh, 2007). Apparently, the farmers had alternatives to remarriage, such as employing servants, to keep the households intact. This situation was the opposite of that pertaining to first marriages, where there were no effects of economic fluctuations among the landless groups, only among the landed groups (Bengtsson, 2014). After years of high food prices, the landed groups delayed the marriage of their children, just as they delayed their leaving home (Dribe, 2000). Thus, marriages and remarriages worked in different ways for the landed and landless groups.

These findings confirm that southern Sweden was part of the Western European family pattern regarding the dominance of nuclear families, a high age at marriage and large proportions of the population not marrying. This pattern was the opposite of the Eastern European family pattern, in which stem-family households dominated and marriage was early and universal among the women.

However, the findings question the saving model put forward by Wrigley and Schofield (1981) as the main reason for the difference in living standards between the East and the West. Research based on the SEDD does not support the idea that during bad times saving enough money to marry and set up a new household took longer, thereby slowing the population growth and promoting living standards. For some people, marriage became possible simply by finding a place where both the man and the woman could work. For those buying a farm, not only their own but also parental savings were important, which is why a two-generation model considering parental wealth is necessary to understand the timing of marriages (Bengtsson, 2014).

## 5 FERTILITY

The theory of the demographic transition stated that modernisation first caused a decline in the mortality rate and, after a delay, a decline in the fertility rate, partly due to a decline in the infant and child mortality rates and partly due to modernisation (Notestein, 1953). After obtaining the number of surviving children they wanted, parents tried to avoid further births using some form of contraception. This situation meant that the decline in the fertility rate mainly should have occurred among women in their late thirties. However, this stopping behaviour is not completely true for Sweden, where the fertility rate declined at all ages above 25, but the older age groups showed a larger decline early in the transition than the younger age groups did (Bengtsson & Ohlsson, 1994; Dribe, 2009). The decline in the fertility rate is of such a magnitude that it rules out the decline in the infant and child mortality rates as a main factor. Moreover, the timing rules out the infant mortality rate as the main explanation since it had already started to decline in the mid-18th century (Bengtsson & Ohlsson, 1994). This finding is consistent with previous research showing that stopping behaviour was only part of the story and that later starting and delaying childbearing were also important, regardless of whether these behaviours were due to an adjustment to the new economic conditions or they were an effect of an innovation process (Carlsson, 1966).

A study using the SEDD found large differences in the start of the fertility decline between different socioeconomic groups, where the elite groups were forerunners, and the unskilled workers, laggards (Bengtsson & Dribe, 2014). The study showed that the fertility transition involved not only parity-specific stopping but also prolonged birth intervals. This finding meant that even newly married couples controlled their fertility, which is consistent with the findings for the entire country (Bengtsson & Ohlsson, 1994). It also showed that the interval between the marriage and the first birth was initially shorter for the lower socioeconomic strata, implying that the marriage and first birth decisions were interlinked in this group. This finding has also been made for other populations and implies that first births need to be analysed separately from second and higher order births. Finally, turning to second and higher order births, the study demonstrated that the elite group and the middle class were the first to start to limit their fertility, followed by the skilled workers and farmers and finally the unskilled workers. Thus, while the fertility rates initially diverged by socioeconomic status, they converged by the 1930s. Overall, a similar pattern has been found for other European and North American populations (Dribe et al., 2017), for Sweden as a whole based on census data (Dribe & Scalone, 2014), and for Stockholm in the same period (Dribe & Molitoris, 2016).

The socioeconomic pattern of the fertility transition in rural Scania does not appear completely consistent with several of the major explanations, such as the infant mortality decline, increased female labour-force participation, and a quantity-quality trade-off. Instead, it is consistent with an innovation process in which new ideas and attitudes about family limitation spread from the elite to other social groups. Whether this explanation for the observed pattern holds true is difficult to ascertain. High benefits of having children and comparatively low costs could also help to explain the lag in the decline among the farmers and labourers, but the early decline of the elite group seems difficult to reconcile with this explanation.

The relationship between social class and fertility in Scania (Landskrona and the five parishes) changed over the 20th century (Dribe & Smith, 2020). Pronounced changes occurred in the associations between social class and fertility over time, which during some periods, depended on parity. A higher position was associated with high fertility for the men and lower fertility for women before 1970, which then converged to a positive association for both sexes after 1990. Over the same timeframe, the

weakly U-shaped relationship between social class and continued childbearing turned into a positive association for second births and a largely negative association for higher order births.

Prior to the fertility transition, the married women in the five rural parishes gave birth to an average of seven children, slightly less than Sweden as a whole (Bengtsson & Dribe, 2010a). The question is whether this finding means that families did not plan their births before the fertility decline. Did the women avoid pregnancies in years of hardship? A clear fertility response to short-term economic stress has been found in studies for Sweden and other countries using aggregated data (Bengtsson & Ohlsson, 1978; Galloway, 1988). The crude birth rate in Sweden followed real wages for the agricultural workers very closely until the late 19th century. In fact, the two series are almost impossible to distinguish (Bengtsson, 2000). A closer look at the lag structure by an analysis of the monthly data revealed that the fertility response began a few months after the harvest and reached its maximum 15 to 17 months afterwards, independent of the marriage rate (Bengtsson & Ohlsson, 1988).

We would expect different socioeconomic groups to respond differently to economic stress for many reasons. While the landed groups had shorter birth intervals than the landless groups did, food prices primarily affected the latter (Bengtsson & Dribe, 2006, 2010a, 2010b; Dribe, 2000). Similar to the results of the analyses using the aggregated data, the response came after a few months and persisted for more than a year. The fact that it came so early indicates that fertility was deliberately controlled since miscarriages typically occurred in the beginning of pregnancy, giving delayed effects on births. The fact that bad harvests could be anticipated also supports this conclusion (Bengtsson & Dribe, 2006). The delayed response, however, might also have been a by-product of miscarriages and sub-fecundity due to acute malnutrition. As real wages increased among the landless groups in the latter part of the 19th century and consumption stabilised, food prices no longer had an impact on the fertility rates (Bengtsson & Dribe, 2005).

Taken together, the studies using the SEDD have shown that deliberate fertility control occurred among the workers prior to the fertility transition, not primarily to limit family size but to alleviate stress and maintain consumption. These studies have also shown that the fertility transition was about not only stopping but also spacing. It started with the upper classes in the 1880s, was followed by the middle classes and farmers, and was concluded by the workers in the 1930s. Finally, no link was found between declining infant and child mortality and fertility, most importantly, since the time gap between the two was very long but also because the fertility decline was much larger than the decline in the infant and child mortality. Hence, these findings question the previous understanding of the fertility transition and the absence of family planning prior to it, showing that many families controlled births to cope with their present economic situations.

## 6 MORTALITY

In Sweden, the mortality rate among infants started to decline in the 18th century, similar to the United States and England (Fogel, 1996; Fridlitzius, 1984; Wrigley, Davies, Oeppen, & Schofield, 1997). Based on aggregate statistical tables for Sweden, which have information on causes of death, it can be concluded that the early decline was entirely due to a reduction in smallpox mortality (Bengtsson, 2001). It was followed in the beginning of the 19th century by a decline in child mortality independent of social and economic structures (Fridlitzius, 1984). Then, starting in the mid-19th century, a decline occurred in adult mortality, which was faster among the women than the men (Bengtsson & Ohlsson, 1994). Changes in the causes of death over the long term followed the epidemiological transition, from a predominance of acute infectious diseases, e.g., epidemics, to chronic and human-caused diseases (Omran, 1971).

Identifying the role of social and economic factors in the mortality rate decline using the SEDD annual data is, however, not an easy task. The main reason is the small number of deaths when considering not only occupation but also age and sex. Instead, most studies using the SEDD have analysed the mortality levels and the response to short-term economic stress for subperiods for the different socioeconomic groups.

Starting with socioeconomic differences in adult mortality rates, the SEDD-based research, including the data for Landskrona from 1905, has shown that, contrary to the established view, the mortality

gradient is quite a recent phenomenon (Bengtsson, Dribe, & Helgertz, 2020; Debiasi, 2020), consistent with our previous findings for the rural and semi-urban parishes (Bengtsson & Dribe, 2011; Bengtsson & van Poppel, 2011). A full mortality gradient in working ages by social class and income emerged only after 1950 for women and after 1970 for men and even later for the elderly people. The estimated differentials for the period after 1990 were very close to those for the entire country (Torssander & Erikson, 2010), as were the results for the beginning of the 20th century (Dribe & Eriksson, 2018). Thus, in 1920–1950, for which the SEDD includes data for both the countryside and the town, adult mortality for men exhibited no socioeconomic differences but possibly some for women. This finding holds also when examining occupations in detail. The men with more prestigious jobs — such as architect, engineer, physician, and lawyer — did not have lower mortality rates than those in other occupations until the second half of the 20th century (Debiasi, 2020).

Advantages related to higher social class, or higher income, appeared at approximately the same time for different disease groups, regardless of preventability, which indicates that the emergence of the gradient was not dependent on medical treatment. An exception was that, already in 1920–1950, a higher income for men was associated with lower mortality from infectious diseases. Another outstanding result was that, during the 19th and first half of the 20th century, the mortality from circulatory diseases for the men had a reversed social gradient (Debiasi & Dribe, 2020).

In the rural area (the five parishes) in 1813–1864, the unskilled adult men and women had a higher mortality than the low-skilled workers did. For men, higher classes had higher mortality than the low-skilled workers did (Bengtsson & Dribe, 2011; Bengtsson, Dribe, & Helgertz, 2020), which created a U-shaped relationship between social class and mortality for men. Similar findings have been made for other areas in Sweden (Edvinsson & Broström, 2012; Edvinsson & Lindkvist, 2011) in the 19th century and for Sweden as a whole in the early 20th century (Dribe & Eriksson, 2018). Apparently, adult mortality differed between the socioeconomic groups prior to the emergence of a complete mortality gradient, whether due to exposure, lifestyles, or other factors. These differences were, however, not systematic, did not form a gradient, and were not always even in an expected direction given the access to resources.

There was no clear pattern of class differentials in the childhood mortality in the first half of the 19th century, a period when both infant and child mortality declined rapidly (Dribe & Karlsson, 2021; Johansson, 2004). However, there were certain differences between areas depending on soil type, which could be viewed as a proxy for agricultural productivity (Hedefalk, 2016; Hedefalk, Quaranta, & Bengtsson, 2017). In the second half of the 19th century, class differences started to emerge. This was the case for both the post-neonatal infant mortality rate and the child mortality rate. Over time, an essentially full gradient was established for post-neonatal mortality, and a weak gradient emerged for child mortality. However, the most striking pattern was the disadvantaged position for the lowest class of the unskilled workers, a disadvantage that remained throughout the 1960s, also at a time when the mortality levels were very low, and the living standards had increased dramatically for all classes in the population (Dribe & Karlsson, 2021).

Before the second half of the 19th century, knowledge was limited on how common infectious diseases were transmitted, except for epidemic diseases such as smallpox and whooping cough. This fact made it difficult even for the upper classes and the best educated to protect their children, and themselves, from disease. When the transmitting agents became known, and the means were available, the lower classes did not lag far behind in their use. This was, for example, the case for smallpox vaccination, which became compulsory in 1816 (Dribe & Nystedt, 2003). This was also the case for infectious diseases in the latter part of the 19th century. When their mode of infection and preventive measures became known, for example, the use of antiseptics and isolation hospitals, all social groups benefitted from it at approximately the same time (Lazuka, 2017, 2018; Lazuka, Quaranta, & Bengtsson, 2016). Similarly, improvements in water and sanitation were explicitly targeted towards the entire population, not to specific social groups (Helgertz & Önnersfors, 2019).

Analyses of the national data at the aggregated level show that fluctuations in real wages had strong impacts on the child and adult mortality rates until the mid-19th century, which implies that at least part of the population lived close to the margins (Bengtsson & Ohlsson, 1985). Nevertheless, the levels of child and adult mortality in this period exhibited no gradient. Does this finding mean that the entire population suffered from increasing mortality as food prices increased, which seems very odd given the inequality in access to land?

The results using the SEDD have shown that, in the beginning of the 19th century, after the agricultural reforms, families with different access to land were affected differently by economic stress (Bengtsson, 2004b; Bengtsson & Dribe, 2000). The working classes suffered from high food prices to such an extent that family members died (Bengtsson, 2000, 2002, 2004b). Their infants were barely affected, possibly due to breastfeeding, nor were the elderly people. Instead, it was adults and children of both sexes that died from high food prices (Bengtsson, 2004b). Before then, in the latter part of the 18th century, all social groups suffered from short-term economic stress. This was a period when epidemics, such as smallpox, whooping cough, and typhoid fever, caused many deaths. The similar mortality experience of rich and poor individuals in this period might have been, at least partly, a result of the spread of such diseases due to temporary migration in years of poor harvests rather than malnutrition. Later on, in the latter part of the 19th century, as agricultural transformation and industrialisation raised real wages for the workers, the workers no longer died in years of economic stress (Bengtsson & Dribe, 2005).

Determining causality empirically is important but difficult. It involves two different, but related, problems with regard to the mortality response to short-term economic stress. The first is the extent to which the impact of food prices on the mortality rate is biased when selecting years with mortality crises, a method used in many studies. In a study using the SEDD data for 1765–1898, Bengtsson and Broström (2011) found that conducting a study that focuses only on mortality crisis years led to a large overestimation of the impact of food prices on the mortality rate. The second problem concerns the mixing of factors that directly and indirectly have an impact on mortality. Using the additive hazards model, in combination with a dynamic path analysis, they demonstrated that, while food prices had an effect on socioeconomic position in adulthood, the direct effect of food prices on old-age mortality was dominant. Their finding is consistent with the late emergence of a social gradient in adult mortality (Bengtsson & Dribe, 2011; Bengtsson, Dribe, & Helgertz, 2020).

These results present contrasting views regarding the role of diet and disease in the mortality transition. On the one hand, the impact of food prices on the mortality rate changed with the development of agriculture and industry. With the agricultural reforms in the beginning of the 19th century, the mortality rate no longer depended on food prices among the farmers, only among the workers. When real wages increased in the latter part of the 19th century, the mortality rate no longer depended on food prices for any socioeconomic group. Thus, the mortality response to fluctuations in food prices had a social gradient, which successively disappeared over the 19th century. On the other hand, there was no social gradient in the mortality level until the second half of the 19th century, when a gradient started to emerge for children. It was followed by women in working ages, then men of working ages, and finally with elderly people. This finding meant that the entire emergence of the social gradient in mortality took place over a one-hundred-year period, starting in the second part of the 19th century, following the same age pattern as the mortality decline. Taken together, these results suggest that socioeconomic status has not always been an important determinant of health and mortality, as has often been argued (e.g., Link & Phelan, 1995).

## 7 A LIFE-COURSE PERSPECTIVE ON HEALTH AND PROSPERITY

Studies of the mortality rate decline in England and Wales as well as Sweden using national data have shown that the children born in the first part of the 19th century who successively experienced lower mortality rates also had lower mortality rates as they aged (Crimmins & Finch, 2006; Fridlitzius, 1989; Kermack, McKendrick, & KcKinley, 1934). This pattern is, however, less clear for children born at the end of the 19th century (Fridlitzius, 1989). The first question is whether delayed or lasting, effects of improved health in early life are an effect of improved nutrition, reduced disease load, medical treatment in childhood, or something else. The second question is whether improved health in childhood was passed on to the next generation (Floud, Fogel, Harris, & Hong, 2011; Fogel & Costa, 1997).

Since these questions are difficult to answer using national data, efforts have been made to use regional data instead. Studies of Norway and England have shown that general living conditions at the time around a person's birth had an impact on heart disease mortality rates at older ages (Barker & Osmond, 1986; Forsdal, 1977). Later analyses have shown that such a correlation existed not only

for heart disease but also for many other diseases, e.g., respiratory and allergic diseases, diabetes, hypertension, breast and testicular cancers, and neuropsychiatric diseases (Kuh, Ben-Shlomo, & Susser, 2004; Lindström & Davey Smith, 2007/2019). While national or regional data help us to describe the time pattern, they do not help us to determine the causes, since many other factors show similar trends. For this reason, attention has been drawn to the individual-level data.

Studies based on the individual-level data for the United States and several other countries, including Sweden, show that exposure to the 1918 pandemic during the foetal stage led to health damage in later life (Almond, 2006; Helgertz & Bengtsson, 2019). Studies of malnutrition during WWII in the Netherlands show similar effects (Lumey, 1998). Studies of a single event, such as a pandemic or a war, have obvious identification problems, simply because other factors may change at the same time. To solve this problem, the SEDD has been used to analyse the long-term effects of repeated events, such as variations in food prices and disease exposure at the start of life.

Given the strong effects of food prices on female adult mortality among the landless in the first part of the 19th century, many pregnant women may have suffered from malnutrition, possibly resulting in health problems for their new-borns. The infant mortality also showed great variability, often due to the spread of non-nutrition-dependent epidemics, such as smallpox and whooping cough, creating variation in the disease load between birth cohorts. Since short-term deviations in food prices and the infant mortality rate were not correlated, we used them as independent exogenous indicators of nutrition and disease exposure in early life.

Bengtsson and Lindström (1997, 2000, 2003), analysing the rural parishes in the SEDD, 1765–1898, found that exposure to epidemics, such as smallpox and whooping cough, in the first years of life had a strong impact on health in adult life. Nutritional deprivation and socioeconomic adversity during the foetal stage or in the first year of life had no such effect. The mortality in infections and heart diseases at older ages were particularly dependent on exposure to infectious diseases in the first year of life (Bengtsson & Lindström, 2000). Effects of disease load in the first year of life have also been found for ages 25–55 (Bengtsson, 1997b). Regarding the mechanisms, Bengtsson and Lindström (2003) suggest that infections in early life hampered the development of organs and cells and caused the onset of inflammation processes that led to arterial sclerosis.

Efforts have also been made to analyse the effects of conditions at the start of life on the mortality rate in childhood and adolescence (see Johansson, 2004). Analysing the impact of food prices and the disease load in ages 1–15 for different socioeconomic groups, Johansson (2004), using the same data as Bengtsson and Lindström, finds that the effects differ over time as well as between different social groups. Claesson (2009), using the data for five rural parishes, found that in 1831, a year with a severe outbreak of whooping cough, new-borns showed higher mortality rates at ages 15–25 than the surrounding birth cohorts. Quaranta (2013, 2014), studying the mortality rates at all ages above age one in the five rural parishes from 1813 to 1968, found that selection dominated in early childhood, followed by a period up to age 25 in which the effects of selection and scarring cancelled out, after which scarring started to dominate. Analysing outbreaks of different epidemics — such as whooping cough, measles, and scarlet fever — Quaranta (2013) found that whooping cough in the first years of life was particularly harmful.

The role of exposure to infectious diseases in the first years of life may affect not only the mortality later in life but also physiological well-being more generally. At the aggregated level, harvests during the foetal stage affected the proportions being dismissed when they were later tested for the army (Hellstenius, 1871). However, using the rural parishes in the SEDD between 1818 and 1868, Öberg (2014a, 2015) did not find any early-life effects on physical status, measured by the height of males at conscription. This finding is consistent with the fact that selection and scarring tended to cancel out until approximately age 25 (Quaranta, 2013, 2014). Overall, the differences in heights between social groups were small. The men whose fathers belonged to the higher socioeconomic groups were slightly taller than average, but the sons of the farmers and farm workers had similar heights (Öberg 2014a, 2014b). For the women, the health of their new-born children has been used as an indicator of their own health, which showed that the women who were exposed to high infant mortality rates, mainly due to epidemics, more often lost their babies in the first year of life than the women who were not exposed to high infant mortality rates (Quaranta, 2013).

Infant mortality was transmitted across generations: mothers who lost two or more of their siblings in infancy had a higher likelihood of experiencing the death of their own offspring in their first year of life (Quaranta, 2018; Quaranta & Sommerseth, 2018). The same pattern was also found in northern Sweden (Broström, Edvinsson, & Engberg, 2018), Belgium (Donrovich, Puschmann, & Matthijs, 2018), the Netherlands (van Dijk & Mandemakers, 2018), and Norway (Sommerseth, 2018).

Conditions in early life potentially affected not only the mortality rate at older ages but also income and socioeconomic position. Therefore, the question is whether the health effect observed was due to scarring that appeared later in life or to the inability to accumulate resources across the life course, which could have prevented mortality in later life (see Bengtsson & Mineau, 2009). Did conditions in early life affect health in later life directly or indirectly? While Bengtsson and Broström (2009), for the rural parishes during the 19th century, found effects of being exposed to diseases in the first year of life on both socioeconomic position at age 50 and mortality at older ages, they found no effects of adult socioeconomic conditions on the mortality rate. This finding means that adverse conditions in early life had a direct effect on the mortality rate in later life. The lack of effects of adult socioeconomic positions is consistent with previous analyses of the adult and old age mortality rates for the same period, which showed the lack of a social gradient (Bengtsson & Dribe, 2011; Bengtsson, Dribe, & Helgertz, 2020; Debiasi, 2020).

As the disease environment in the first years of life had such a strong impact on health later in life, one would expect successful public health interventions that reduced infant mortality rates to have had lasting health and economic impact as well. Exactly this finding was made in a series of studies on the introduction of medically trained midwives and isolation hospitals at the end of the 19th century using a causal approach (Lazuka, 2018, 2019, 2020; Lazuka, Quaranta, & Bengtsson, 2016). Similar results were found for the entire country using the individual-level data from 1968 onwards (Lazuka, 2019). The effects were universal and somewhat stronger among individuals from poor socioeconomic backgrounds and at higher baseline levels of disease burden. The fact that interventions directed at stopping the spread of infectious disease had such success strengthens the conclusion that the effects of variations in infectious diseases in the first year of life on later-life health were causal.

The weak effects of parental resources, measured by socioeconomic status, on health in later life for children born in the 18th and 19th centuries (Bengtsson & Lindström, 1997, 2000, 2003; Quaranta, 2013, 2014) are consistent with findings showing that the social gradient in the infant and child mortality rates emerged only in the latter part of the 19th century (Dribe & Karlsson, 2021). It is also consistent with our finding that the social gradient emerged much later for the adults and elderly individuals (Bengtsson & Dribe, 2011; Bengtsson, Dribe, & Helgertz, 2020).

In a study using the SEDD, Bengtsson and Broström (2008) explored the role played by inheritance for longevity by estimating a model of the overall mortality rate among married persons aged 50 years and above, considering genetic as well as socioeconomic and environmental factors. They considered whether these factors had temporary or long-lasting effects on health. They found that the age at death of the mother and the father had persistent impacts on their adult children's overall mortality regardless of sex, even after controlling for early-life factors and socioeconomic and environmental factors throughout the life course. In addition, they found strong birth cohort effects and effects of the disease load in the first year of life on male offspring but were unable to find any effects of socioeconomic status, either at the time of birth or achieved later in life, a result consistent with earlier findings (Bengtsson & Broström, 2008; see also Bengtsson & Mineau, 2008). Birth cohort factors seemed to become weaker, relatively speaking, during the 20th century, as indicated both by the aggregated and individual-level studies for Sweden (Fridlitzius, 1989; Lindström, 2015).

Other life-course experiences may also have long-term health impacts. A much-discussed issue is the extent to which childbearing has repercussions for women's health and mortality later in life, as suggested by evolutionary demographic theories about a trade-off between reproduction and longevity (e.g., Westendorp & Kirkwood, 1998). The number of children born reduced post-reproductive longevity for the women in Scania but not for the men (Dribe, 2004b). More importantly, only the landless women showed this association, which was interpreted as support for social and economic mechanisms rather than as evolutionary trade-offs. A similar conclusion was reached in a comparative study, including the SEDD data, which showed that post-reproductive longevity was reduced by having more children, especially for the women who were widowed at a young age (Alter, Dribe, & van Poppel, 2007). This

finding points to the social and economic circumstances under which children are born and reared as crucial for the long-term mortality effects.

## 8 THE LONG ROAD TO HEALTH AND PROSPERITY — A SUMMARY

The escape from a society characterised by hunger and premature death to a prosperous welfare state was an incremental process. Since the public poor law system provided help only to a few disabled and elderly people, if people could not provide for themselves, most of them had to rely on their families. The pension system, which was introduced in 1913, was not sufficient, and did not cover the most basic costs of living until 1948. Agricultural development in the beginning of the 19th century, together with the trade liberation of farm products, improved the living conditions for the farmers, regardless of whether they owned or rented the land. They used their own savings and altered their household sizes and compositions to relieve them of short-term economic stress caused by variations in harvests and the prices of their products. The workers who lacked these opportunities tried to cope with the economic stress by delaying births instead. Nevertheless, they suffered from increased mortality rates when food prices increased. They shared the burden equally between men and women, children and adults. It was not until the latter part of the 19th century, when job opportunities within industrialisation opened and real wages grew that their food consumption became stable enough on a yearly basis to not affect their deaths.

The improvements in living conditions for the landed, but not the landless, that came with agricultural development were not reflected in the mortality levels. Instead, a social gradient in mortality rates emerged only in the latter part of the 19th century, first for children, then, from the mid-20th century onwards, also for men and women in working ages, and finally for the elderly people. The late emergence of a social gradient in mortality implies that factors other than nutrition were even more crucial.

While food prices had an almost instant effect on mortality among the landless, until the latter part of the 19th century, we find no long-term effects. Instead, exposure to diseases, such as smallpox and whooping cough, in the first years of life had lasting health effects. Since such scarring was often cancelled by selection at younger ages, such improvements did not always appear until later in life. Public health interventions to stop the spread of infectious diseases not only had instant but also lasting effects on individuals, underlining the importance of improving the health conditions for indigent people and children.

Although the landless groups delayed childbirth to cope with short-term economic stress, they were the last group to limit family size. Instead, in the 1880s, the higher classes started to limit family size, followed by the middle class and the farmers. Finally, in the 1920s, the workers started to control family size. Apparently, ways of controlling fertility were known among the workers for a long time, since they delayed births in times of economic stress although not to control family size.

The inheritance law of 1857, which gave equal inheritance to all children regardless of sex, changed the situation for the middle classes and farmers. Having many children made it difficult to transfer enough resources to safeguard the children's economic standing. Thus, children would be less able to support their parents as they retired. Since the workers typically did not leave many resources for their children when they retired, they lacked this motivation. Instead, they benefited from the wage income of their children.

After a period of proletarianisation, the development of the industrial sector and urbanisation opened new opportunities for the working class. Later, in the post-industrial period, the demand for education played a similar role. At that time, people planned their families not to stabilise consumption but to combine work and family life. While no one suffered from hunger any longer and access to health care became practically free, social differences in the mortality rate continued to increase more rapidly than ever before.

## ACKNOWLEDGEMENTS

This paper has been produced within the research programme 'Landskrona Population Study', funded by the Swedish Foundation for Humanities and Social Sciences (RJ). SEDD is part of SwedPop which is a national infrastructure funded by The Swedish Research Council (VR) together with the partner institutions, Umeå University, Lund University, University of Gothenburg, the national Archives, and the Stockholm City Archives.

We are grateful to Christer Lundh and Luciana Quaranta for comments and suggestions.

## REFERENCES

- Allen, R. C., Bengtsson, T., & Dribe M. (2005). Introduction. In R. C. Allen, T. Bengtsson, & M. Dribe (Eds.), *Living standards in the past: New perspectives on well-being in Asia and Europe* (pp. 1–22). Oxford: Oxford University Press. Available from [10.1093/0199280681.003.0001](https://doi.org/10.1093/0199280681.003.0001)
- Almond, D. (2006). Is the 1918 influenza pandemic over? Long-term effects of in utero influenza exposure in the post-1940 U.S. population. *Journal of Political Economy*, 114(4), 672–712. doi: [10.1086/507154](https://doi.org/10.1086/507154)
- Alter, G., Dribe, M., & van Poppel, F. (2007). Widowhood, family size, and post-reproductive mortality: A comparative analysis of three populations in nineteenth-century Europe. *Demography*, 44(4), 785–806. doi: [10.1353/dem.2007.0037](https://doi.org/10.1353/dem.2007.0037)
- Barker, D. J. P., & Osmond, C. (1986). Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. *The Lancet*, 327(8489), 1077–1081. doi: [10.1016/S0140-6736\(86\)91340-1](https://doi.org/10.1016/S0140-6736(86)91340-1)
- Bengtsson, T. (1987). Migration och löner. Tillämpning av Todaros migrationsteori på 1800-talets svenska urbanisering. In O. Krantz & R. Ohlsson (Eds.), *Ekonomisk-historiska vingslag* (pp. 21–38). Lund: Ekonomisk-historiska institutionen.
- Bengtsson, T. (1990). Migration, wages, and urbanization in Sweden in the nineteenth century. In A. D. van der Woude, A. Hayami, & J. de Vries (Eds.), *Urbanization in history: A process of dynamic interactions* (pp. 186–204). Oxford: Clarendon Press.
- Bengtsson, T. (1997a). Combined life event and time series analysis: The impact of economic fluctuations and household cycles on mortality in rural Sweden, 1750–1850. *EurAsian Project on Population and Family History Working Paper Series* (No. 5). Kyoto: International Research Center for Japanese Studies.
- Bengtsson, T. (1997b). Adult mortality in Sweden, 1760–1895. *EurAsian Project on Population and Family History Working Paper Series* (No. 8). Kyoto: International Research Center for Japanese Studies.
- Bengtsson, T. (1999). The vulnerable child. Economic insecurity and child mortality in pre-industrial Sweden: A case study of Västanafors, 1757–1850. *European Journal of Population*, 15, 117–151. doi: [10.1023/A:1006215701608](https://doi.org/10.1023/A:1006215701608)
- Bengtsson, T. (2000). Inequality in death: Effects of the agrarian revolution in southern Sweden, 1765–1865. In T. Bengtsson & O. Saito (Eds.), *Population and economy. From hunger to modern economic growth* (pp. 301–333). Oxford: Oxford University Press.
- Bengtsson, T. (2001). Mortality: The great historical decline. In N. J. Smelser & P. B. Yates (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 10079–10085). Oxford: Pergamon. doi: [10.1016/B0-08-043076-7/02116-1](https://doi.org/10.1016/B0-08-043076-7/02116-1)
- Bengtsson, T. (2002). Why dad dies. The mortality of men in their working ages in the 18th and 19th centuries. In R. Derosas & M. Oris (Eds.), *When dad died. Individuals and families coping with distress in past societies* (pp. 81–98). Bern: Peter Lang. Available from <https://www.peterlang.com/view/title/60571>
- Bengtsson, T. (2004a). Living standards and economic stress. In T. Bengtsson, C. Campbell, & J. Z. Lee (Eds.), *Life under pressure: Mortality and living standards in Europe and Asia, 1700–1900* (pp. 27–59). Cambridge, MA: MIT Press. doi: [10.7551/mitpress/4227.003.0007](https://doi.org/10.7551/mitpress/4227.003.0007)
- Bengtsson, T. (2004b). Mortality and social class in four Scanian parishes, 1765–1865. In T. Bengtsson, C. Campbell, & J. Z. Lee (Eds.), *Life under pressure: Mortality and living standards in Europe and Asia, 1700–1900* (pp. 135–172). Cambridge, MA: MIT Press. doi: [10.7551/mitpress/4227.003.0012](https://doi.org/10.7551/mitpress/4227.003.0012)

- Bengtsson, T. (2014). The influence of economic factors on first marriages in historical Europe and Asia. In C. Lundh & S. Kurosu (Eds.), *Similarity in difference: Marriage in Europe and Asia, 1700–1900* (pp. 121–168). Cambridge, MA: MIT Press. doi: [10.7551/mitpress/9944.003.0011](https://doi.org/10.7551/mitpress/9944.003.0011)
- Bengtsson, T., & Broström, G. (2008). Inheritance, environment, and mortality in older ages, southern Sweden, 1813–1894. In T. Bengtsson & G. P. Mineau (Eds.), *Kinship and demographic behavior in the past. International studies in population* (Vol. 7, pp. 185–201). Dordrecht: Springer. doi: [10.1007/978-1-4020-6733-4\\_8](https://doi.org/10.1007/978-1-4020-6733-4_8)
- Bengtsson, T., & Broström, G. (2009). Do conditions in early life affect old-age mortality directly and indirectly? Evidence from 19th-century rural Sweden. *Social Science & Medicine* 68(9), 1583–1590. doi: [10.1016/j.socscimed.2009.02.020](https://doi.org/10.1016/j.socscimed.2009.02.020)
- Bengtsson, T., & Broström, G. (2010). Mortality crises in rural southern Sweden 1766–1860. In S. Kurosu, T. Bengtsson & C. Campbell (Eds.), *Demographic responses to economic and environmental crises* (pp. 1–16). Kashiwa: Reitaku University Press. Retrieved from <https://www.iussp.org/sites/default/files/AllArticles.pdf>
- Bengtsson, T., & Broström, G. (2011). Famines and mortality crises in 18th to 19th century southern Sweden. *GENUS*, 67(2), 119–139. Retrieved from <http://www.jstor.org/stable/genus.67.2.119>
- Bengtsson, T., & Dribe, M. (1997). Economy and demography in western Scania, Sweden, 1650–1900. *EurAsian Project on Population and Family History Working Paper Series* (No. 10). Kyoto: International Research Center for Japanese Studies.
- Bengtsson, T., & Dribe, M. (2000). Risk management in the family. Demographic responses to short-term economic stress in southern Sweden during the agricultural transformation, 1829–1865. In M. Neven & C. Capron (Eds.), *Family structures, demography and population. A comparison of societies in Asia and Europe* (pp. 159–184). Liège: Laboratory of Demography at the University of Liège.
- Bengtsson, T., & Dribe, M. (2005). New evidence on the standard of living in Sweden during the eighteenth and nineteenth centuries: Long-term development of the demographic response to short-term economic stress. In R. C. Allen, T. Bengtsson, & M. Dribe (Eds.), *Living standards in the past: New perspectives on well-being in Asia and Europe* (pp. 341–371). Oxford: Oxford University Press. doi: [10.1093/0199280681.003.0015](https://doi.org/10.1093/0199280681.003.0015)
- Bengtsson, T., & Dribe, M. (2006). Deliberate control in a natural fertility population: Southern Sweden 1766–1864. *Demography*, 43(4), 727–746. doi: [10.1353/dem.2006.0030](https://doi.org/10.1353/dem.2006.0030)
- Bengtsson, T., & Dribe, M. (2010a). Agency, social class, and fertility in southern Sweden 1766–1865. In N. O. Tsuya, F. Wang, G. Alter, & J. Z. Lee (Eds.), *Prudence and pressure: Reproduction and human agency in Europe and Asia, 1700–1900* (pp. 159–194). Cambridge, MA: MIT Press. doi: [10.7551/mitpress/8162.003.0014](https://doi.org/10.7551/mitpress/8162.003.0014)
- Bengtsson, T., & Dribe, M. (2010b). Economic stress and reproductive responses. In N. O. Tsuya, F. Wang, G. Alter, & J. Z. Lee (Eds.), *Prudence and pressure: Reproduction and human agency in Europe and Asia, 1700–1900* (pp. 97–127). Cambridge, MA: MIT Press. doi: [10.7551/mitpress/8162.003.0011](https://doi.org/10.7551/mitpress/8162.003.0011)
- Bengtsson, T., & Dribe, M. (2011). The late emergence of socioeconomic mortality differentials: A micro-level study of adult mortality in southern Sweden 1815–1968. *Explorations in Economic History*, 48(3), 389–400. doi: [10.1016/j.eeh.2011.05.005](https://doi.org/10.1016/j.eeh.2011.05.005)
- Bengtsson, T., & Dribe, M. (2014). The historical fertility transition at the micro level: Southern Sweden 1815–1939. *Demographic Research*, 30, 493–534. doi: [10.4054/DemRes.2014.30.17](https://doi.org/10.4054/DemRes.2014.30.17)
- Bengtsson, T., Dribe, M., & Helgertz, J. (2020). When did the health gradient emerge? Social class and adult mortality in southern Sweden, 1813–2015. *Demography*, 57(3), 953–977. doi: [10.1007/s13524-020-00877-5](https://doi.org/10.1007/s13524-020-00877-5)
- Bengtsson, T., & Jörberg, L. (1975). Market integration in Sweden during the 18th and 19th centuries. Spectral analysis of grain prices. *Economy and History*, 18(2), 93–106. doi: [10.1080/00708852.1975.10418922](https://doi.org/10.1080/00708852.1975.10418922)
- Bengtsson, T., & Lindström, M. (1997). Old age mortality by cause of death in rural Scania, 1760–1894. *EurAsian Project on Population and Family History Working Paper Series* (No. 9). Kyoto: International Research Center for Japanese Studies
- Bengtsson, T., & Lindström, M. (2000). Childhood misery and disease in later life: The effects on mortality in old-age of hazards experienced in early life, southern Sweden, 1760–1894. *Population Studies*, 54(3), 263–277. doi: [10.1080/713779096](https://doi.org/10.1080/713779096)
- Bengtsson, T., & Lindström, M. (2003). Airborne infectious diseases during infancy and mortality in later life, southern Sweden, 1766–1894. *International Journal of Epidemiology*, 32(2), 286–294. doi: [10.1093/ije/dyg061](https://doi.org/10.1093/ije/dyg061)

- Bengtsson, T., & Mineau, G. P. (Eds.). (2008). *Kinship and demographic behavior in the past. International studies in population* (Vol. 7). Dordrecht: Springer. doi: [10.1007/978-1-4020-6733-4](https://doi.org/10.1007/978-1-4020-6733-4)
- Bengtsson, T., & Mineau, G. P. (2009). Introduction. Early-life effects on socio-economic performance and mortality in later life: A full life-course approach using contemporary and historical sources. *Social Science & Medicine*, 68(9), 1561–1564. doi: [10.1016/j.socscimed.2009.02.012](https://doi.org/10.1016/j.socscimed.2009.02.012)
- Bengtsson, T., & Oeppen, J. (1993). A reconstruction of the population of Scania, 1650–1760. *Lund Papers in Economic History* (No. 32). Lund: Lund University, Department of Economic History. Retrieved from [https://portal.research.lu.se/portal/files/89673885/LUP\\_1993\\_32.pdf](https://portal.research.lu.se/portal/files/89673885/LUP_1993_32.pdf)
- Bengtsson, T., & Ohlsson, R. (1978). Befolkning och konjunkturer. *Meddelande från Ekonomisk-historiska institutionen* (No. 1). Lund: Lund University.
- Bengtsson, T., & Ohlsson, R. (1984). Population and economic fluctuations: Sweden 1749–1869. In T. Bengtsson, G. Fridlitzius, & R. Ohlsson (Eds.), *Pre-industrial population change: The mortality decline and short-term population movements* (pp. 277–298). Lund: Almqvist & Wiksell International.
- Bengtsson, T., & Ohlsson, R. (1985). Age-specific mortality and short-term changes in the standard of living: Sweden, 1751–1859. *European Journal of Population*, 1(4), 309–326. doi: [10.1007/BF01797146](https://doi.org/10.1007/BF01797146)
- Bengtsson, T., & Ohlsson, R. (1988). Seasonality in births and economic cycles: Sweden 1749–1869. *Meddelande från Ekonomisk-historiska institutionen* (No. 56). Lund: Lund University.
- Bengtsson, T., & Ohlsson, R. (1994). The demographic transition revised. In T. Bengtsson (Ed.), *Population, economy, and welfare in Sweden* (pp. 13–36). Heidelberg/Berlin: Springer. doi: [10.1007/978-3-642-85170-4\\_2](https://doi.org/10.1007/978-3-642-85170-4_2)
- Bengtsson, T., & van Poppel, F. (2011). Socioeconomic inequalities in death from past to present: An introduction. *Explorations in Economic History*, 48(3), 343–356. doi: [10.1016/j.eeh.2011.05.004](https://doi.org/10.1016/j.eeh.2011.05.004)
- Berkner, L. (1972). The stem family and the developmental cycle of the peasant household: An eighteenth-century Austrian example. *The American Historical Review*, 77(2), 398–418. doi: [10.2307/1868698](https://doi.org/10.2307/1868698)
- Broström, G., Edvinsson, S., & Engberg, E. (2018). Intergenerational transfers of infant mortality in 19th-century northern Sweden. *Historical Life Course Studies*, 7, 106–122. doi: [10.51964/hlcs9282](https://doi.org/10.51964/hlcs9282)
- Carlsson, G. (1966). The decline of fertility: Innovation or adjustment process. *Population Studies*, 20(2), 147–174. doi: [10.1080/00324728.1966.10406092](https://doi.org/10.1080/00324728.1966.10406092)
- Claësson, O. (2009). *Geographical differences in infant and child mortality during the initial mortality decline. Evidence from southern Sweden, 1749–1830* (Licentiate thesis, Lund University). Lund: Lund University.
- Crafts, N., & Mills, T. C. (2020). The race between population and technology: Real wages in the first Industrial Revolution. *CAGE online working papers* (No. 502). Warwick: The University of Warwick. Retrieved from <https://warwick.ac.uk/fac/soc/economics/research/centres/cage/manage/publications/wp502.2020.pdf>
- Crimmins, E. M., & Finch, C. E. (2006). Infection, inflammation, height, and longevity. *PNAS*, 103(2), 498–503. doi: [10.1073/pnas.0501470103](https://doi.org/10.1073/pnas.0501470103)
- de Vries, J. (1994). The industrial revolution and the industrious revolution. *Journal of Economic History*, 54(2), 249–270. Retrieved from <http://www.jstor.org/stable/2123912>
- Debiasi, E. (2020). *The historical origins of the mortality gradient: Socioeconomic inequalities in adult mortality over two centuries in Sweden* (Doctoral dissertation, Lund University). Lund: Media-Tryck. Retrieved from [https://portal.research.lu.se/portal/en/publications/the-historical-origins-of-the-mortality-gradient\(4727723a-de82-42f8-ac47-15608dc151f1\).html](https://portal.research.lu.se/portal/en/publications/the-historical-origins-of-the-mortality-gradient(4727723a-de82-42f8-ac47-15608dc151f1).html)
- Debiasi, E., & Dribe, M. (2020). SES inequalities in cause-specific adult mortality: A study of the long-term trends using longitudinal individual data for Sweden (1813–2014). *European Journal of Epidemiology*, 35, 1043–1056. doi: [10.1007/s10654-020-00685-6](https://doi.org/10.1007/s10654-020-00685-6)
- Donrovich, R., Puschmann, P., & Matthijs, K. (2018). Mortality clustering in the family: Fast life history trajectories and the intergenerational transfer of infant death in late 19th- and early 20th-century Antwerp, Belgium. *Historical Life Course Studies*, 7, 47–68. doi: [10.51964/hlcs9285](https://doi.org/10.51964/hlcs9285)
- Dribe, M. (2000). *Leaving home in a peasant society: Economic fluctuations, household dynamics and youth migration in southern Sweden, 1829–1866*. Södertälje: Almqvist & Wiksell International.
- Dribe, M. (2003a). *Liv och rörelse. Familj och flyttningar i 1800-talets svenska bondesamhälle*. Hedemora: Gidlunds.

- Dribe, M. (2003b). Migration of rural families in nineteenth century southern Sweden. A longitudinal analysis of local migration patterns. *The History of the Family*, 8(2), 247–265. doi: [10.1016/S1081-602X\(03\)00028-9](https://doi.org/10.1016/S1081-602X(03)00028-9)
- Dribe, M. (2003c). Dealing with economic stress through migration: Lessons from nineteenth century rural Sweden. *European Review of Economic History*, 7(3), 271–299. doi: [10.1017/S1361491603000108](https://doi.org/10.1017/S1361491603000108)
- Dribe, M. (2004a). Leaving home as a family strategy in times of economic and demographic stress: The study of rural Scania, Sweden 1829–1866. In F. van Poppel, M. Oris, & J. Lee (Eds.), *The road to independence: Leaving home in European and Asian Societies* (pp. 85–116). Bern: Peter Lang.
- Dribe, M. (2004b). Long-term effects of childbearing on mortality: Evidence from pre-industrial Sweden. *Population Studies*, 58(3), 297–310. doi: [10.1080/0032472042000272357](https://doi.org/10.1080/0032472042000272357)
- Dribe, M. (2009). Demand and supply factors in the fertility transition: A county-level analysis of age-specific marital fertility in Sweden, 1880–1930. *European Review of Economic History*, 13(1), 65–94. doi: [10.1017/S1361491608002372](https://doi.org/10.1017/S1361491608002372)
- Dribe, M., Breschi, M., Gagnon, A., Gauvreau, D., Hanson, H. A., Maloney, T. N., Mazzone, S., Molitoris, J., Pozzi, L., Smith, K. R., & Vézina, H. (2017). Socioeconomic status and fertility decline: Insights from historical transitions in Europe and North America. *Population Studies*, 71(1), 3–21. doi: [10.1080/00324728.2016.1253857](https://doi.org/10.1080/00324728.2016.1253857)
- Dribe, M., & Eriksson, B. (2018). Socioeconomic status and adult life expectancy in early 20th-century Sweden: Evidence from full-count micro census data. *Lund Papers in Economic Demography* (No. 1). Lund: Lund University. Retrieved from [https://www.ed.lu.se/media/ed/papers/working\\_papers/LPED%202018%201.pdf](https://www.ed.lu.se/media/ed/papers/working_papers/LPED%202018%201.pdf)
- Dribe, M., & Helgertz, J. (2016). The lasting impact of grandfathers: Class, occupational status, and earnings over three generations in Sweden 1815–2011. *Journal of Economic History*, 76(4), 969–1000. doi: [10.1017/S0022050716000991](https://doi.org/10.1017/S0022050716000991)
- Dribe, M., Helgertz, J., & Van de Putte, B. (2015). Did social mobility increase during the industrialization process? A micro-level study of a transforming community in southern Sweden 1828–1968. *Research in Social Stratification and Mobility*, 41(1), 25–39. doi: [10.1016/j.rssm.2015.04.005](https://doi.org/10.1016/j.rssm.2015.04.005)
- Dribe, M., & Karlsson, O. (2021). Inequality in early life: Social class differences in childhood mortality in southern Sweden, 1815–1967. *Economic History Review*, Online first. doi: [10.1111/ehr.13089](https://doi.org/10.1111/ehr.13089)
- Dribe, M., & Lundh, C. (2005a). Gender aspects of inheritance strategies and land transmission in rural Scania, Sweden, 1720–1840. *The History of the Family*, 10(3), 293–308. doi: [10.1016/j.hisfam.2005.03.005](https://doi.org/10.1016/j.hisfam.2005.03.005)
- Dribe, M., & Lundh, C. (2005b). Retirement as a strategy for land transmission: A micro study of pre-industrial rural Sweden. *Continuity and Change*, 20(2), 165–191. doi: [10.1017/S0268416005005497](https://doi.org/10.1017/S0268416005005497)
- Dribe, M., & Lundh, C. (2005c). People on the move: Determinants of servant migration in nineteenth-century Sweden. *Continuity and Change*, 20(1), 53–91. doi: [10.1017/S026841600400534X](https://doi.org/10.1017/S026841600400534X)
- Dribe, M., & Lundh, C. (2005d). Finding the right partner: Rural homogamy in nineteenth-century Sweden. *International Review of Social History*, 50(S13), 149–177. doi: [10.1017/S0020859005002105](https://doi.org/10.1017/S0020859005002105)
- Dribe, M., & Lundh, C. (2009a). Partner choice and intergenerational occupational mobility: The case of nineteenth-century rural Sweden. *Continuity and Change*, 24(3), 487–512. doi: [10.1017/S0268416009990178](https://doi.org/10.1017/S0268416009990178)
- Dribe, M., & Lundh, C. (2009b). Status homogamy in the preindustrial marriage market: Partner selection according to age, social origin, and place of birth in nineteenth-century rural Sweden. *Journal of Family History*, 34(4), 387–406. doi: [10.1177/0363199009344708](https://doi.org/10.1177/0363199009344708)
- Dribe, M., & Lundh, C. (2010). Marriage choices and social reproduction: The interrelationship between partner selection and intergenerational socioeconomic mobility in 19th-century Sweden. *Demographic Research*, 22, 347–382. doi: [10.4054/DemRes.2010.22.14](https://doi.org/10.4054/DemRes.2010.22.14)
- Dribe, M., & Lundh, C. (2014). Social norms and human agency: Marriage in nineteenth-century Sweden. In C. Lundh & S. Kurosu (Eds.), *Similarity in difference: Marriage in Europe and Asia, 1700–1900* (pp. 211–260). Cambridge, MA: MIT Press. doi: [10.7551/mitpress/9944.003.0014](https://doi.org/10.7551/mitpress/9944.003.0014)
- Dribe, M., Lundh, C., & Nystedt, P. (2007). Widowhood strategies on preindustrial society. *Journal of Interdisciplinary History*, 38(2), 207–232. doi: [10.1162/jinh.2007.38.2.207](https://doi.org/10.1162/jinh.2007.38.2.207)
- Dribe, M., Manfredini, M., & Oris, M. (2014). The roads to reproduction: Comparing life-course trajectories in preindustrial Eurasia. In C. Lundh & S. Kurosu (Eds.), *Similarity in difference. Marriage in Europe and Asia, 1700–1900* (pp. 89–120). Cambridge, MA: MIT Press. doi: [10.7551/mitpress/9780262027946.003.0004](https://doi.org/10.7551/mitpress/9780262027946.003.0004)

- Dribe, M., & Molitoris, J. (2016). Ready to stop: Socioeconomic status and the fertility transition in Stockholm, 1878–1926. *Economic History Review*, 69(2), 679–704. doi: [10.1111/ehr.12275](https://doi.org/10.1111/ehr.12275)
- Dribe, M., & Nystedt, P. (2003). Information, trust and diffusion of smallpox vaccination. The case of Scania, Sweden 1802–1835. *Scandinavian Economic History Review*, 51(1), 9–28. doi: [10.1080/03585522.2003.10410835](https://doi.org/10.1080/03585522.2003.10410835)
- Dribe, M., Olsson, M., & Svensson, P. (2010). Economic crisis, manorialism, and demographic response: Southern Sweden in the preindustrial period. In S. Kurosu, T. Bengtsson, & C. Campbell (Eds.), *Demographic responses to economic and environmental crises* (pp. 17–47). Kashiwa: Reitaku University Press. Retrieved from <https://www.iussp.org/sites/default/files/AllArticles.pdf>
- Dribe, M., Olsson, M., & Svensson, P. (2017). The agricultural revolution and the conditions of the rural poor, southern Sweden 1750–1860. *Economic History Review*, 70(2), 483–508. doi: [10.1111/ehr.12378](https://doi.org/10.1111/ehr.12378)
- Dribe, M., & Quaranta, L. (2020). The Scanian Economic-Demographic Database (SEDD). *Historical Life Course Studies*, Online first. doi: [10.51964/hlcs9302](https://doi.org/10.51964/hlcs9302)
- Dribe, M., & Scalone, F. (2014). Social class and net fertility before, during, and after the demographic transition: A micro-level analysis of Sweden 1880–1970. *Demographic Research*, 30, 429–464. doi: [DemRes.2014.30.15](https://doi.org/10.18632/demres.2014.30.15)
- Dribe, M., & Smith, C. D. (2020). Social class and fertility: A long-run analysis of southern Sweden, 1922–2015. *Population Studies*, Online publication. doi: [10.1080/00324728.2020.1810746](https://doi.org/10.1080/00324728.2020.1810746)
- Dribe, M., & Svensson, P. (2008a). Social mobility in nineteenth century rural Sweden: A micro level analysis. *Scandinavian Economic History Review*, 56(2), 122–141. doi: [10.1080/03585520802137194](https://doi.org/10.1080/03585520802137194)
- Dribe, M., & Svensson, P. (2008b). Statarnas geografiska rörlighet, sociala ursprung och fruktsamhet. In C. Lundh & M. Olsson (Eds.), *Statarliv — i myt och verklighet* (pp. 187–210). Hedemora: Gidlunds.
- Dribe, M., & Svensson, P. (2019). Landskrona 1900–2000 — A comparative analysis of the economic and demographic development. *Lund Papers in Economic Demography* (No. 3). Lund: Lund University. Retrieved from [https://www.ed.lu.se/media/ed/papers/working\\_papers/LPED\\_2019\\_3.pdf](https://www.ed.lu.se/media/ed/papers/working_papers/LPED_2019_3.pdf)
- Dribe, M., & Van de Putte, B. (2012). Marriage seasonality and the industrious revolution: Southern Sweden 1690–1895. *Economic History Review*, 65(3), 1123–1146. Retrieved from <http://www.jstor.org/stable/23271565>
- Edvinsson, S., & Broström, G. (2012). Old age, health and social inequality: Exploring the social patterns of mortality in 19th century northern Sweden. *Demographic Research*, 26, 633–330. doi: [10.4054/DemRes.2012.26.23](https://doi.org/10.4054/DemRes.2012.26.23)
- Edvinsson, S., & Lindkvist, M. (2011). Wealth and health in 19th century Sweden. A study of social differences in adult mortality in the Sundsvall Region. *Explorations in Economic History*, 48(3), 376–388. doi: [10.1016/j.eeh.2011.05.007](https://doi.org/10.1016/j.eeh.2011.05.007)
- Elmér, Å. (1971). *Från Fattigsverige till välfärdsstaten: Sociala förhållanden och socialpolitik i Sverige under nittonhundratalet*. Lund: Berlings.
- Erikson, R., & Goldthorpe, J. H. (1992). *The constant flux: A study of class mobility in industrial societies*. Oxford: Clarendon Press.
- Featherman, D. L., Jones, F. L., & Hauser, R. M. (1975). Assumptions of social mobility research in the U.S.: The case of occupational status. *Social Science Research*, 4(4), 329–360. doi: [10.1016/0049-089X\(75\)90002-2](https://doi.org/10.1016/0049-089X(75)90002-2)
- Floud, R., Fogel, R. W., Harris, B., & Hong, S. C. (2011). *The changing body: Health, nutrition and human development in the Western world since 1700*. Cambridge: Cambridge University Press. doi: [10.1017/CBO9780511975912](https://doi.org/10.1017/CBO9780511975912)
- Fogel, R. W. (1996). *The escape from hunger and premature death 1700–2100: Europe, America and the Third World*. The 1996 Ellen McArthur Lectures. Cambridge: Cambridge University Press. Retrieved from <http://ndl.ethernet.edu.et/bitstream/123456789/56684/1/118.pdf>
- Fogel, R. W., & Costa, D. L. (1997). A theory of technophysio evolution, with some implications for forecasting population, health care costs, and pension costs. *Demography*, 34(1), 49–66. doi: [10.2307/2061659](https://doi.org/10.2307/2061659)
- Forsdal, A. (1977). Are poor living conditions in childhood and adolescence an important risk factor for arteriosclerotic heart disease? *British Journal of Preventive and Social Medicine*, 31(2), 91–95. doi: [10.1136%2Fjebch.31.2.91](https://doi.org/10.1136%2Fjebch.31.2.91)
- Fridlitzius, G. (1984). The mortality decline in the first phase of the demographic transition: Swedish experiences. In T. Bengtsson, G. Fridlitzius & R. Ohlsson (Eds.), *Pre-industrial population change: The mortality decline and short-term population movements* (pp. 71–114). Lund: Almqvist & Wiksell International.

- Fridlitzius, G. (1989). The deformation of cohorts: Nineteenth century mortality decline in a generational perspective. *Scandinavian Economic History Review*, 37(3), 3–17. doi: [10.1080/03585522.1989.10408152](https://doi.org/10.1080/03585522.1989.10408152)
- Galloway, P. R. (1988). Basic patterns in annual variation in fertility, nuptiality, mortality, and prices in pre-industrial Europe. *Population Studies*, 42(2), 275–303. doi: [10.1080/0032472031000143366](https://doi.org/10.1080/0032472031000143366)
- Grusky, D. B., & Hauser, R. M. (1984). Comparative social mobility revisited: Models of convergence and divergence in 16 countries. *American Sociological Review*, 49(1), 19–38. doi: [10.2307/2095555](https://doi.org/10.2307/2095555)
- Hajnal, J. (1965). European marriage patterns in perspective. In D. V. Glass & D. E. C. Eversley (Eds.), *Population in history. Essays in historical demography* (pp. 101–144). London: Edward Arnold.
- Hajnal, J. (1983). Two kinds of pre-industrial formation systems. In R. Wall, J. Robin, & P. Laslett (Eds.), *Family forms in historic Europe* (pp. 65–104). Cambridge: Cambridge University Press. doi: [10.1017/CBO9780511897535.003](https://doi.org/10.1017/CBO9780511897535.003)
- Hammel, E. A., & Laslett, P. (1974). Comparing household structure over time and between cultures. *Comparative Studies in Society and History*, 16(1), 73–109. doi: [10.1017/S0010417500007362](https://doi.org/10.1017/S0010417500007362)
- Heckscher, E. (1949). *Sveriges ekonomiska historia* (Vol. 2.1). Stockholm: Albert Bonniers Förlag.
- Hedefalk, F. (2016). *Life paths through space and time: Adding the micro-level geographic context to longitudinal historical demographic research* (Doctoral dissertation, Lund University). doi: [10.13140/RG.2.2.14187.13605](https://doi.org/10.13140/RG.2.2.14187.13605)
- Hedefalk, F., & Dribe, M. (2020). The social context of nearest neighbors shapes educational attainment regardless of class origin. *Proceedings of the National Academy of Sciences*, 117(26), 14918–14925. doi: [10.1073/pnas.1922532117](https://doi.org/10.1073/pnas.1922532117)
- Hedefalk, F., Quaranta, L., & Bengtsson, T. (2017). Unequal lands: Soil type, nutrition and child mortality in southern Sweden, 1850–1914. *Demographic Research*, 36, 1039–1080. doi: [10.4054/DemRes.2017.36.36](https://doi.org/10.4054/DemRes.2017.36.36)
- Helgertz, J., & Bengtsson, T. (2019). The long-lasting influenza: The impact of fetal stress during the 1918 influenza pandemic on socioeconomic attainment and health in Sweden, 1968–2012. *Demography*, 56(4), 1389–1425. doi: [10.1007/s13524-019-00799-x](https://doi.org/10.1007/s13524-019-00799-x)
- Helgertz, J., Bengtsson, T., & Dribe, M. (2020). Income data in the Scanian Economic Demographic Database (SEDD). *Lund Papers in Economic Demography* (No. 6). Lund: Lund University. Retrieved from [https://www.ed.lu.se/media/ed/papers/working\\_papers/LPED\\_2020\\_6.pdf](https://www.ed.lu.se/media/ed/papers/working_papers/LPED_2020_6.pdf)
- Helgertz, J., & Önerfors, M. (2019). Public water and sewerage investments and the urban mortality decline in Sweden 1875–1930. *The History of the Family*, 24(2), 307–338. doi: [10.1080/1081602X.2018.1558411](https://doi.org/10.1080/1081602X.2018.1558411)
- Hellstenius, J. (1871). Skördarna i Sverige och deras verkningar. *Statistisk Tidskrift*, 29, 77–119.
- Hochstadt, S. (1999). *Mobility and modernity: Migration in Germany, 1820–1989*. Ann Arbor, MI: University of Michigan Press.
- Johansson, K. (2004). *Child mortality during the demographic transition. A longitudinal analysis of a rural population in southern Sweden, 1766–1894* (Doctoral dissertation, Lund University). Lund: Almqvist & Wiksell International.
- Kermack, W. O., McKendrick, A. G., & McKinley P. L. (1934). Death rates in Great Britain and Sweden. Some general regularities and their significance. *International Journal of Epidemiology*, 30(4), 678–683. doi: [10.1093/ije/30.4.678](https://doi.org/10.1093/ije/30.4.678)
- Kuh, D., Ben-Shlomo, Y., & Ezra, S. (Eds.). (2004). *A life course approach to chronic disease epidemiology*. Oxford: Oxford University Press. Available from <https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780198578154.001.0001/acprof-9780198578154>
- Kurosu, S., Lundh, C., & Breschi, M. (2014). Remarriage, gender, and rural households: A comparative analysis of widows and widowers in Europe and Asia. In C. Lundh & S. Kurosu (Eds.), *Similarity in difference: Marriage in Europe and Asia, 1700–1900* (pp. 169–208). Cambridge, MA: MIT Press. doi: [10.7551/mitpress/9944.003.0012](https://doi.org/10.7551/mitpress/9944.003.0012)
- Lambert, P. S., Zijdeman, R. L., van Leeuwen, M. H. D., Maas, I., & Prandy, K. (2013). The construction of HISCAM: A stratification scale based on social interactions for historical comparative research. *Historical Methods*, 46(2), 77–89. doi: [10.1080/01615440.2012.715569](https://doi.org/10.1080/01615440.2012.715569)
- Laslett, P. (1965/1994). *The world we have lost: Further explored*. London: Routledge.
- Laslett, P., & Wall, R. (Eds.). (1972). *Households and families in past time*. Cambridge: Cambridge University Press. doi: [10.1017/CBO9780511561207](https://doi.org/10.1017/CBO9780511561207)
- Lazuka, V. (2017). *Defeating disease: Lasting effects of public health and medical breakthroughs between 1880 and 1945 on health and income in Sweden* (Doctoral dissertation, Lund University). Lund: Media-Tryck.

- Lazuka, V. (2018). The long-term health benefits of receiving treatment from qualified midwives at birth. *Journal of Development Economics*, 133, 415–433. doi: [10.1016/j.jdeveco.2018.03.007](https://doi.org/10.1016/j.jdeveco.2018.03.007)
- Lazuka, V. (2019). Early-life assets in oldest-old age: Evidence from primary care reform in early twentieth century Sweden. *Demography*, 56(2), 679–706. doi: [10.1007/s13524-018-0758-4](https://doi.org/10.1007/s13524-018-0758-4)
- Lazuka, V. (2020). Infant health and later-life labour market outcomes: Evidence from the introduction of sulpha antibiotics in Sweden. *Journal of Human Resources*, 55(2), 660–698. doi: [10.3368/jhr.55.3.0817-9016R](https://doi.org/10.3368/jhr.55.3.0817-9016R)
- Lazuka, V., Quaranta, L., & Bengtsson, T. (2016). Fighting infectious disease: Evidence from Sweden 1870–1940. *Population and Development Review*, 42(1), 27–52. doi: [10.1111/j.1728-4457.2016.00108.x](https://doi.org/10.1111/j.1728-4457.2016.00108.x)
- Lee, R. D. (1981). Short-term variation: Vital rates, prices, and weather. In E. A. Wrigley & R. S. Schofield, *The population history of England, 1541–1871* (pp. 356–401). Cambridge: Cambridge University Press.
- Lindström, M. (2015). *All that's mine I carry with me. Early life disease and adult health in Sweden during 250 years.* (Doctoral dissertation, Lund University). Lund: Media-Tryck.
- Lindström, M., & Davey Smith, G. (2007/2019). A life course perspective to the modern secular mortality decline and socioeconomic differences in morbidity and mortality in Sweden. In T. Bengtsson & N. Keilman (Eds.), *Old and new perspectives on mortality forecasting*, Demographic Research Monographs (pp. 295–309). Cham: Springer. doi: [10.1007/978-3-030-05075-7\\_21](https://doi.org/10.1007/978-3-030-05075-7_21)
- Link, B. G., & Phelan, J. (1995). Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior*, 80–94. doi: [10.2307/2626958](https://doi.org/10.2307/2626958)
- Lipset, S. M., & Zetterberg, H. (1956). A theory of social mobility. *Transactions of the Third World Congress in Sociology* (Vol. 3). London: International Sociological Association.
- Lumey, L. H. (1998). Reproductive outcomes in women prenatally exposed to undernutrition: A review of findings from the Dutch famine birth cohort. *Proceedings of the Nutrition Society*, 57, 129–135. doi: [10.1079/pns19980019](https://doi.org/10.1079/pns19980019).
- Lundh, C. (1993). Giftermålsmonster i Sverige före det industriella genombrottet. *Lund Papers in Economic History* (No. 31). Lund: Lund University, Department of Economic History.
- Lundh, C. (1995). Households and families in pre-industrial Sweden. *Continuity and Change*, 10(1), 33–68. doi: [10.1017/S0268416000002496](https://doi.org/10.1017/S0268416000002496)
- Lundh, C. (1999a). The social mobility of servants in rural Sweden, 1740–1894. *Continuity and Change*, 14(1), 57–89. doi: [10.1017/S0268416099003264](https://doi.org/10.1017/S0268416099003264)
- Lundh, C. (1999b). Servant migration in Sweden in the early nineteenth century. *Journal of Family History*, 24(1), 53–73. doi: [10.1177%2F036319909902400104](https://doi.org/10.1177%2F036319909902400104)
- Lundh, C. (1999c). Marriage and economic change in Sweden during the 18th and 19th centuries. In I. Devos I & L. Kennedy (Eds.), *Marriage and rural economy. Western Europe since 1400* (pp. 217–241). Turnhout: Brepols. doi: [10.1484/M.CORN-EB.4.00130](https://doi.org/10.1484/M.CORN-EB.4.00130)
- Lundh, C. (2002). Remarriages in Sweden in the 18th and 19th centuries. *History of the Family*, 7(3), 423–449. doi: [10.1016/S1081-602X\(02\)00112-4](https://doi.org/10.1016/S1081-602X(02)00112-4)
- Lundh, C. (2007). Remarriage, gender and social class: A longitudinal study of remarriage in southern Sweden, 1766–1894. *Continuity and Change*, 22(3), 373–406. doi: [10.1017/S0268416007006443](https://doi.org/10.1017/S0268416007006443)
- Lundh, C. (2013). The geography of marriage: Regional variations in age at first marriage in Sweden 1870–1900. *Scandinavian Journal of History*, 38(3), 318–343. doi: [10.1080/03468755.2013.804003](https://doi.org/10.1080/03468755.2013.804003)
- Lundh, C., & Kurosu, S. (Eds.). (2014). *Similarity in difference: Marriage in Europe and Asia, 1700–1900*. Cambridge, MA: MIT Press. doi: [10.7551/mitpress/9780262027946.001.0001](https://doi.org/10.7551/mitpress/9780262027946.001.0001)
- Lundh, C., & Olsson, M. (2002). The institution of retirement on Scanian estates in the nineteenth century. *Continuity and Change*, 17(3), 373–403. doi: [10.1017/S0268416002004393](https://doi.org/10.1017/S0268416002004393)
- Lundh, C., & Olsson, M. (2009). Tenancy contracts in Scania from the Middle Ages to the nineteenth Century. In B. van Bavel & P. Schofield (Eds.), *The development of leasehold in Northwestern Europe, c. 1200–1600*, Comparative rural history of the North Sea area, No. 10 (pp. 113–137). Turnhout: Brepols. doi: [10.1484/M.CORN-EB.3.317](https://doi.org/10.1484/M.CORN-EB.3.317)
- Malthus, T. R. (1803). *An essay on the principle of population* (2nd ed.). Cambridge: Cambridge University Press. Retrieved from [https://assets.cambridge.org/97805214/19543/frontmatter/9780521419543\\_frontmatter.pdf](https://assets.cambridge.org/97805214/19543/frontmatter/9780521419543_frontmatter.pdf)
- Notestein, F. W. (1953). Economic problems of population change. In *Proceedings of the Eighth International Conference of Agricultural Economists* (pp. 13–31). London: Oxford University Press.

- Nystedt, P. (2002). Widowhood-related mortality in Scania during the 19th century. *History of the Family*, 7, 451–478. doi: [10.1016/S1081-602X\(02\)00113-6](https://doi.org/10.1016/S1081-602X(02)00113-6)
- Öberg, S. (2014a). *Social bodies: Family and community level influences on height and weight, southern Sweden 1818–1968* (Doctoral dissertation, University of Gothenburg). Gothenburg: University of Gothenburg. Retrieved from [https://gupea.ub.gu.se/bitstream/2077/35270/1/gupea\\_2077\\_35270\\_1.pdf](https://gupea.ub.gu.se/bitstream/2077/35270/1/gupea_2077_35270_1.pdf)
- Öberg, S. (2014b). Long-term changes of socioeconomic differences in height among young adult men in southern Sweden, 1818–1968. *Economics & Human Biology*, 15, 140–152. doi: [10.1016/j.ehb.2014.08.003](https://doi.org/10.1016/j.ehb.2014.08.003)
- Öberg, S. (2015). The direct effect of exposure to disease in early life on the height of young adult men in southern Sweden, 1814–1948. *Population Studies*, 69(2), 179–199. doi: [10.1080/00324728.2015.1045545](https://doi.org/10.1080/00324728.2015.1045545)
- Olsson, M., & Svensson, P. (2010). Agricultural growth and institutions: Sweden 1700–1860. *European Review of Economic History*, 14(2), 275–304. doi: [10.1017/S1361491610000067](https://doi.org/10.1017/S1361491610000067)
- Omran, A. R. (1971). The epidemiologic transition: A theory of the epidemiology of population change. *The Milbank Quarterly*, 83(4), 731–57, doi: [10.1111/j.1468-0009.2005.00398.x](https://doi.org/10.1111/j.1468-0009.2005.00398.x)
- Quaranta, L. (2013). *Scarred for life. How conditions in early life affect socioeconomic status, reproduction and mortality in Southern Sweden, 1813–1968* (Doctoral dissertation, Lund University). Lund: Media-Tryck. Retrieved from <https://lup.lub.lu.se/search/ws/files/3026458/3737380.pdf>
- Quaranta, L. (2014). Early life effects across the life course: The impact of individually defined exogenous measures of disease exposure on mortality by sex in 19th- and 20th-century southern Sweden. *Social Science & Medicine*, 119, 266–273. doi: [10.1016/j.socscimed.2014.04.007](https://doi.org/10.1016/j.socscimed.2014.04.007)
- Quaranta, L. (2018). Intergenerational transfers in infant mortality in southern Sweden, 1740–1968. *Historical Life Course Studies*, 7, 88–105. doi: [10.51964/hlcs9283](https://doi.org/10.51964/hlcs9283)
- Quaranta, L., & Sommerseth, H. L. (2018). Introduction: Intergenerational transmissions of infant mortality using the Intermediate Data Structure (IDS). *Historical Life Course Studies*, 7, 1–10. doi: [10.51964/hlcs9288](https://doi.org/10.51964/hlcs9288)
- Skoglund, A.-M. (1992). *Fattigvården på den svenska landsbygden år 1829* (Doctoral dissertation, Stockholm University). Stockholm: Stockholm University.
- Sommerseth, H. L. (2018). The intergenerational transfer of infant mortality in northern Norway during the 19th and early 20th centuries. *Historical Life Course Studies*, 7, 69–87. doi: [10.51964/hlcs9284](https://doi.org/10.51964/hlcs9284)
- Stanfors, M. (2007). *Mellan arbete och familj. Ett dilemma för kvinnor i 1900-talets Sverige*. Stockholm: SNS.
- Svensson, P. (2013). Land market, property rights, and agricultural transformation in southern Sweden, 1680–1870. In G. Béaur, P. R. Schofield, J.-M. Chevet, & M.-T. Perez-Picazo (Eds.), *Property rights, land markets and economic growth in the European countryside (13th–20th centuries)* (pp. 455–473). Turnhout: Brepols. doi: [10.1484/M.RURHE-EB.4.00159](https://doi.org/10.1484/M.RURHE-EB.4.00159)
- Torssander, J., & Eriksson, R. (2010). Stratification and mortality: A comparison of education, class, status, and income. *European Sociological Review*, 26(4), 465–474. doi: [10.1093/esr/jcp034](https://doi.org/10.1093/esr/jcp034)
- Utterström, G. (1957). *Jordbrukets arbetare: Levnadsvillkor och arbetsliv på landsbygden från frihetstiden till mitten av 1800-talet* (Doctoral dissertation, Stockholm College). Stockholm: Tiden.
- Van de Putte, B., & Miles, A. (2005). A social classification scheme for historical occupation data: Partner selection and industrialism in Belgium and England, 1800–1918. *Historical Methods*, 38(2), 61–92. doi: [10.3200/HMTS.38.2.61-94](https://doi.org/10.3200/HMTS.38.2.61-94)
- Van de Putte, B., & Svensson, P. (2010). Measuring social structure in a rural context. Applying the SOCPO-scheme to Scania, Sweden (17th–20th centuries). *Belgisch Tijdschrift voor Nieuwste Geschiedenis*, 40(1–2), 249–293. Retrieved from [10\\_vandeputte\\_svensson.pdf](https://doi.org/10_vandeputte_svensson.pdf) ([journalbelgianhistory.be](http://journalbelgianhistory.be))
- van Dijk, I., & Mandemakers, K. (2018). Like mother, like daughter. Intergenerational transmission of infant mortality clustering in Zeeland, the Netherlands, 1833–1912. *Historical Life Course Studies*, 7, 28–46. doi: [10.51964/hlcs9286](https://doi.org/10.51964/hlcs9286)
- van Leeuwen, M. H. D., & Maas, I. (2011). *HISCLASS: A Historical International Social Class Scheme*. Leuven: Leuven University Press.
- van Leeuwen, M. H. D., Maas, I., & Miles, A. (2002). *HISCO: Historical International Standard Classification of Occupations*. Leuven: Leuven University Press.

- Wargentin, P. (1772/1976). De l'homme; mortalité de l'homme en Suède, comparée à celle de la femme. Du nombre des naissances et des morts dans tous de mois de l'année. In *Mémoires (abrégés) de l'Académie Royale de Sciences de Stockholm*. Paris.
- Westendorp, R. G. J., & Kirkwood, T. B. L. (1998). Human longevity at the cost of reproductive success. *Nature*, 396, 743–746. doi: [10.1038/25519](https://doi.org/10.1038/25519)
- Weibull, C. G. (1923). *Skånska jordbrukets historia. Intill 1800-talets början*. Lund: Gleerups.
- Wrigley, E. A. (2011). Coping with rapid population growth: How England fared in the century preceding the great exhibition of 1851. In D. Feldman & J. Lawrence (Eds.), *Structures and transformations in modern British history* (pp. 24–53). Cambridge: Cambridge University Press.
- Wrigley, E. A., Davies, R. S., Oeppen, J. E., & Schofield, R. S. (1997). *English population history from family reconstitution 1580–1837*. Cambridge: Cambridge University Press. doi: [10.1017/CBO9780511660344](https://doi.org/10.1017/CBO9780511660344)
- Wrigley, E. A., & Schofield, R. S. (1981). *The population history of England 1541–1871*. Cambridge: Cambridge University Press.



HISTORICAL LIFE COURSE STUDIES  
VOLUME 12 (2022), published 01-03-2022

# The Impact of Microdata in Norwegian Historiography 1970 to 2020

Hilde L. Sommerseth

UiT The Arctic University of Norway

Gunnar Thorvaldsen

Ural Federal University & UiT The Arctic University of Norway

## ABSTRACT

The establishment of the Norwegian Historical Data Centre, the 1801 project at the University of Bergen and the data transcriptions and scanned versions of the sources in the National Archives made Norwegian microdata much more available. A more detailed description of the digital techniques applied to the wealth of censuses, church records and other types of nominative data from the 18th century onwards, will be presented in a separate article. Our main focus here is to summarize the impact of the research that has been produced based on the Norwegian historical microdata. These studies span a wide range of fields within social history and historical demography: Emigration, immigration, internal migration, fertility, nuptiality, family history and last but not least mortality studies with a priority given to infant mortality. A recent development is the building of a national historical population register covering the 19th and 20th centuries.

**Keywords:** Norway, Microdata, Censuses, Vital records, Transcription, Migration, Social mobility, Fertility, Mortality, Nuptiality, Family history, Ethnicity, Population register

DOI article: <https://doi.org/10.51964/hlcs11675>

© 2022, Sommerseth, Thorvaldsen

This open-access work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/), which permits use, reproduction & distribution in any medium for non-commercial purposes, provided the original author(s) and source are given credit. See <http://creativecommons.org/licenses/>.

## 1 INTRODUCTION

This article sketches the background of the establishment of the Norwegian Historical Data Centre as part of UiT The Arctic University of Norway. This centre concentrates on data transcriptions and enhanced digital versions of the sources, which have often been developed in cooperation with other database vendors, especially the National Archives. We plan a detailed description of the digital techniques and methods in a separate article. Our focus here is to summarize the impact of the research that has been produced based on the Norwegian historical microdata. The historiographic bottom line is that Norwegian archives abound in censuses, church records and other types of structured individual level data from the 18th century onwards. The advent of the computer from the 1970s onwards helped release these riches and put them at the disposal of students and staff eager to write the population and social history, mostly for local communities or regions (Hubbard, 1995). Since then, we have researched most branches of social and population history, with priority given to studies of mortality, migration and family history. A mixture of methods has been used, for instance fertility has been studied nationally on the municipality level and parish-wise with microdata. While initially researchers used each source separately or with family reconstitution, they are now being linked on the individual level with database techniques. The Norwegian Historical Population Register covering the period from 1800 onwards, is the fruition of this development.

## 2 BACKGROUND

### 2.1 THE FIRST COMPUTER PROJECTS ON INDIVIDUAL LEVEL DATA

Norway's demographic history 1735–1865 was already written in the 1960s by the English demographer Michael Drake (1969) based on aggregates. Still, due to the focus on political history in Norwegian historiography, we could expect continuity with respect to sources and methods from political to social and demographic history. The leader of the first comprehensive social science history research project, Sivert Langholm, had studied the voters in the capital during the 19th century. And two of the project's central master theses studied the members of the radical Thrane protest movement around 1850 and the Workers' Society [Kristiania Arbeidersamfund], respectively. This project, Norwegian social development 1860 to 1900, started in 1971 in the Department of History at the University of Oslo as a well-funded research and teaching project. It aimed to study 19th century Norwegian social history through a micro historical research program. Thus, it promoted historical data on the individual level, facilitating aggregate results on multiple analysis levels, from individuals and families to the inhabitants of census tracts or whole municipalities. The two contextual locations chosen were the capital Oslo, at the time called Kristiania, and rural Ullensaker, a municipality fifty kilometres further northwest (Langholm, 1974, 1975).

The project developed software for transcribing the local sources, correcting errors, listing and sorting the material. In the HISO software package, developed by future National Archivist Ivar Fønnes, there were also opportunities to translate from full-text to encoded version with numerical codes for the different individual characteristics. The results, however, were made available in aggregated form with standard statistical software. The sources treated in this way were the local censuses of 1865 and 1875, parts of the emigrant books, and for rural Ullensaker the church registers 1845 to 1875. The biggest source was the 1875 census for Kristiania with 78,000 individuals. In order to defend the high cost of transferring the sources to computers, the project needed many participants, and over 20 students were recruited to write master theses based on the computerized source material. In addition, a number of students were inspired to address related topics for other parts of the country, including family reconstitutions. Both the University of Bergen and of Oslo were inspired by French demographic research especially Louis Henry's method to link church records for the reconstitution of families (Dyrvik, 1983; Sogner, 2016). Central to the topics addressed were social and geographical mobility. Sølvi Sogner (1979) used her reconstitution of rural Rendalen parish to trace outmigration from the relatively isolated valley towards the coast and later on also to the Netherlands and its colonies, employing marriage banns registers from the Amsterdam City Archives (Sogner, 2012). Today we know significantly more about fertility, nuptiality, mortality and family history, about migrants who moved to America and to the cities, about intergenerational social mobility and about typical careers in the 19th century.

A more advanced use of history oriented computer technology took place in the Department of History at the University of Bergen. Here a project headed by Jan Oldervoll, made the full count census from 1801 with nearly 900,000 records machine-readable and encoded and aggregated the data in cooperation with Statistics Norway (Statistics Norway, 1980). Two master theses were based on record linkage between the 1801 census and the church registers in the surrounding years; cf. Haavet (1982) in the fertility section 6.2 and Engelsen (1983) in the mortality section 4.3. The linking was semi-automatic with automatic normalization of personal names. However, all decisions about which records belonged to the same person were made by manual intervention. Thus, already in the late 1970s they utilized the opportunities inherent in computers equipped with disk storage and interactive terminals. The new techniques facilitated studying far more individuals; 44 randomly selected parishes with a total of 116,000 people. Thus, the pioneer microdata projects in Norway transcribed the full count 1801 census and the late 19th century censuses and some vital records for the capital and a nearby municipality. These were used to study migration, mortality and a range of other topics within the fields of social and population history.

## 2.2 THE NORWEGIAN HISTORICAL DATA CENTRE

The initiative to establish the Norwegian Historical Data Centre (NHDC) was taken by historians at the University of Tromsø (UiT) in 1976. The Demographic Database (DDB) at Umeå University served as a model, where in 1972 professor of pedagogy and former priest, Egil Johansson, kick-started work on transcribing church books in Haparanda, also originally an employment measure (Edvinsson & Engberg, 2020; Thorvaldsen, 1998). Data consultant Jan Olav Hauge prepared the first planning document after visiting Umeå University (Thorvaldsen, 1977, 1978). The plans became concrete against the backdrop of the closure of the last remaining manual telephone exchanges, where many women were losing their employment. Due to political pressure from local and regional authorities, including municipality mayors, the labour market authorities, the County Governor and the UiT, funds were allocated for a pilot project in which two telephone operators transcribed census and church records under the direction of one of the undersigned (Thorvaldsen, 1979). Thus, a professional basis for the NHDC was established, but financing a permanent undertaking turned out to be difficult. But in 1981, during the Parliament's work on the national fiscal budget, three representatives supposed to support the government, broke ranks and voted for the creation of the Norwegian Historical Data Centre together with the opposition. The UiT received economic support for the establishment of two new units, a management unit located in Tromsø, and a unit for transcription and proofreading 150 kilometres to the southeast.

Although the transcription scope of the NHDC was national, targeting at central nominative sources, the initial regional mission of the UiT dictated a focus on source material from the neighbouring region. This was in demand in demographic and social history research and the material was available at the Regional Archives in Tromsø. Still, a choice of period and type of source material had to be made. The period was limited by the Law of Statistics, which rules that nominative material collected by the state is for statistical use only during the first hundred years (<https://www.ssb.no/omssb/lover-og-prinsipper/statistikkloven>). Since the legal authority was initially granted by Parliament in 1907, the censuses from 1865, 1875 and 1900 were available anyway, but the 1910 census was blocked until 2010, when the NHDC and the National Archives made transcribed and encoded versions available. The 1891 census was difficult to handle due to the large amount of person sheets that only existed in the original in the National Archives in Oslo. Three censuses thus became a first priority, also because they were in demand both in quantitatively oriented research and in order to trace historical persons. Church records were blocked for eighty years, but in some cases are difficult to use for transcription because there are backlogs of transfer to the regional archives (Thorvaldsen, 1996). Consequently, the three available national censuses and the regional lists of baptisms, marriages and burials from the church books for Troms province was the focus when starting the work in earnest by 1982. The digital versions were encoded and standardized for statistical use and published as printed editions for genealogical use in the relevant municipalities and parishes.

The National Archives allowed volunteering genealogists to access the scanned version of the 1920 census and start transcriptions while promising to share no information before 1 December 2020. The 1891 census was completed in India with most expenses paid by US companies catering to genealogists: Ancestry, My Heritage and FamilySearch. This is part of a deal negotiated by the National Archives, where the said companies also transcribed the church records for the period from 1815 until about 1930. The 1960 and later censuses were transcribed as part of the contemporary aggregation and went into the building of the Central Population Register which covers the period from 1964

onwards. At the same time, there is much interest in source materials particularly for medical research. We thus need to bridge the time-gap in the mid-20th century. Foremost among these are the 1950 and the 1930 housing and population censuses. These have been scanned by the National Archives as part of our effort to build a Historical Population Register for Norway.

To sum up, by the coordinated efforts of the National Archives and the Norwegian Historical Data Centre, the nominative and full count 1801, 1865, 1875, 1891, 1900 and 1910 censuses have been transcribed, while the censuses from 1920 will be searchable in early 2022. Today, researchers can apply for specific data extraction via our web page, and encoded versions of the censuses until 1910 are also available at the Minnesota Population Center through the North Atlantic Population Project and the Digital Archive of Norway. Our strategic goal for the future is to develop systems for online data extraction and analysis both for our cross-sectional data and our longitudinal register.

### 3 EMIGRATION, IMMIGRATION AND INTERNAL MIGRATION STUDIES

Until the microdata revolution from the 1970s, historical research on migration mainly dealt with emigration overseas, but then internal migration became the focus of many studies. At the start of the millennium, immigration received its fair share of the historians' attention. The literature on migration in Norway is too vast, even if we only consider research based on microdata, to be summarized as part of a historiographic article. Therefore, we shall concentrate on specific examples within the fields of internal migration, emigration and immigration, while also citing references to works providing broader overviews where they exist. For a first attempt to summarize Norwegian domestic migration, see (Thorvaldsen, 2019).

#### 3.1 INTERNAL MIGRATION INSIDE NORWAY

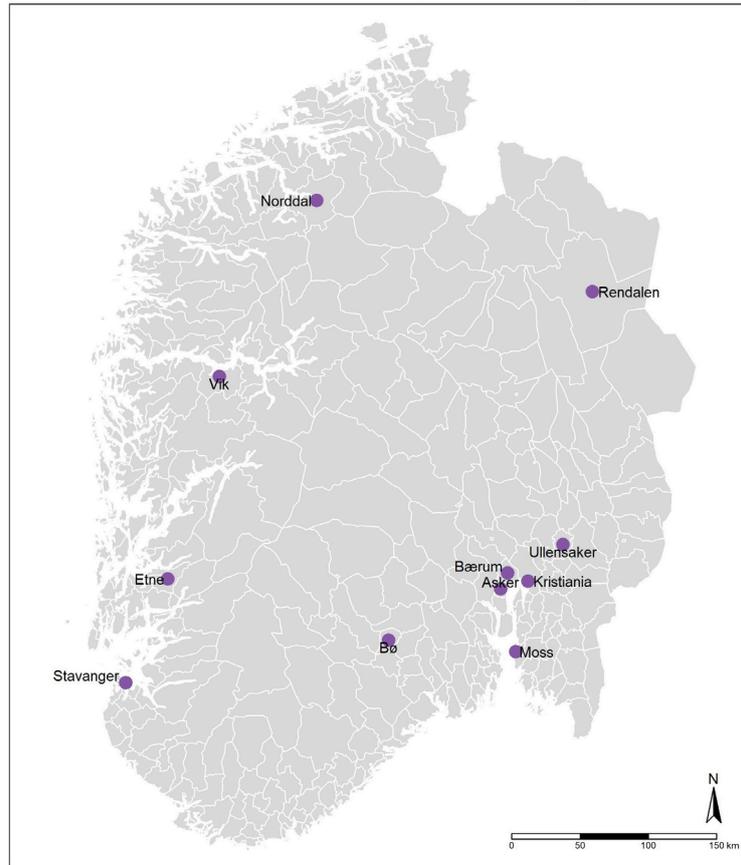
An extensive dissertation by Thorvaldsen (1995) dedicated to *internal* migration, mapped geographical mobility in Troms province in Northern Norway during the second half of the 19th century, when net in-migration turned into net out-migration. The 1865, 1875 and 1900 censuses were used cross-sectionally and longitudinally to estimate in- and out-migration as well as migration between and inside the municipalities. Then it describes what characterized the internal and the in- and out-migrants. The dissertation studied the development of an entire province ("fylke", sometimes called county) based on individual level data, out-migrants from the municipalities being traced to their new domicile. Software for automatic record linkage was developed and used to trace the migrants longitudinally. The amount of in-migration declined during the period 1865–1900, out-migration from the province increased and internal migration stayed at the same level. The empirical results are summarized in a migration model based on the concept of the frontier. The agricultural crisis in the late 1860's and the urban trade crisis around 1880 ended the province's frontier position and the mass in-migration from other parts of the Nordic countries.

Emigration and other out-migration were small in this region because of several barriers to migration. One barrier was the peasant economy, the combination of fishing and farming absorbing the growing population into the family production units. There was no enclosure movement blocking the peasant families from using the resources in the commons as was the case in the British countryside, forcing many to leave for the cities. A second barrier was ethnic, the Sami tended to stay within the areas they dominated, even when they migrated. A third barrier was social. The farmers and peasants, most of them also fishing, tended to not migrate, explaining why this province had the lowest emigration rates in Norway. The exception was the farming community with roots in southern Norway, who emigrated since they had social contacts in the US. They were a minority of inland farmers who had fewer relationships with the rest of the population in the province. Thus, the multiplier effect on emigration inherent in contiguous social networks (Åkerman, 1978), only marginally showed its potential in this part of thinly populated, topographically divided and ethnically mixed Northern Norway. As expected, people mostly migrated to conserve their old way of life. For the fishing peasants of Troms, it was difficult to keep up this combination of trades if they moved to town. Agriculture was marginal to the north on the coast of Finnmark, there was little room for new farms to the south and small chances to fish on the American prairie. Migration and non-migration were to a high degree related to the economy, peasants tending to stay put, while people in the trades and the services migrated more. The social network was a secondary determinant, helping people to decide about their specific destination (Thorvaldsen, 1995).

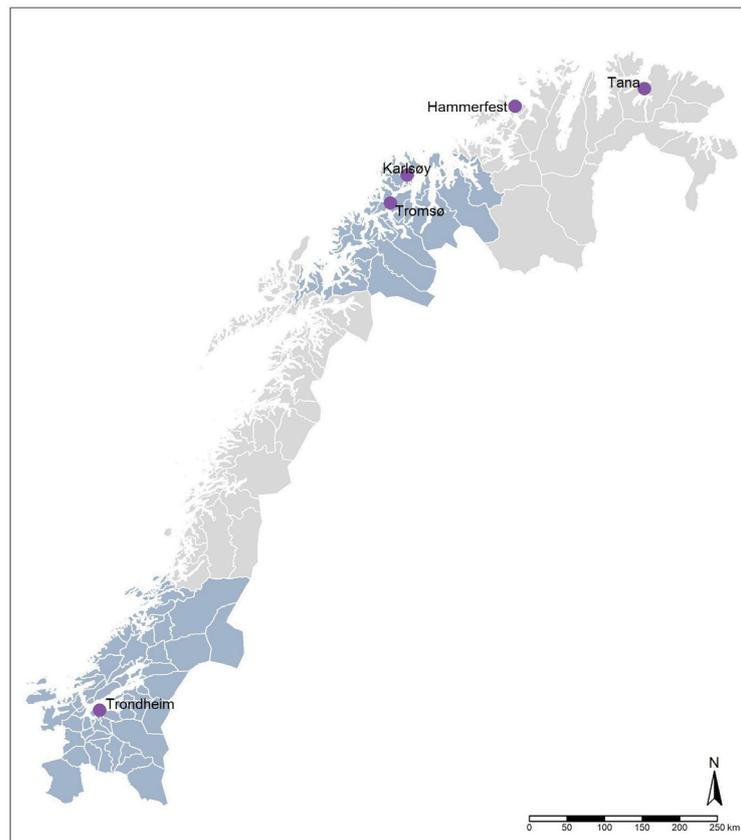
Map 1

Map of Norway

WEST-, EAST- AND SOUTHERN NORWAY



CENTRAL AND NORTHERN NORWAY



Note: The light blue polygons surrounding Trondheim and Tromsø refer to the province specific studies of Trøndelag and Troms respectively. Other place names refer to the location of parish specific studies.

The aim of a large cohort study in mid-Norway (Trøndelag) was to contribute knowledge about both the social and geographic mobility of a relatively oversized cohort born in the mid-19th century, by linking individual level data from church books, censuses and the emigration list database of the National Archives. Lindbekk (2017) followed the birth cohort of 4942 persons from 1855 in Trøndelag (plus in-migrants) until the census of 1910, as well as their parents from 1855 to the 1865 census. As in the Troms study, the municipality divisions of 1855 were used throughout. The mobility was studied against a backdrop of individual birthplaces and parental social backgrounds. On the one hand, there was social equality with respect to child mortality and a significant degree of equal social chances and possibilities for migration. On the other hand, the individual operated at the family's own risk, when many young persons had to break up from familiar relationships, occupations and localities in order to seek "permanent and adequate livelihoods" (Lindbekk, 2017).

### 3.2 EMIGRATION FROM NORWAY

Emigration was studied as part of a study of the varied demographic developments in the region of Trøndelag during the period 1866–1914 (Lindbekk & Opach, 2020), creating comparative municipality maps based on the population censuses. This shows interconnectivity between emigration and internal migration with internal migration as the most extensive. Similar to the national level, the background for mass emigration was the fall in death rates while fertility rates remained stable. Among large, young cohorts only some had prospects for sustainable livelihoods without leaving. Thus, the mobile workforce of young people in search of livelihoods at other places in Norway or America, was a precondition for economic readjustment during the modernizing process.

Rasmus Sunde's doctoral dissertation (2001) about the emigration from Vik parish by the Sognefjord to the US Mid-West is a representative case study about transplanting a specific local European rural population. The dissertation is a quantitative analysis based on reconstructed genealogies of 3319 emigrants, and the database size increases to 12,732 persons when including spouses and children. Using the transcribed church records and the emigration list database of the National Archives with the family reconstitution method, Sunde transformed the genealogical data into measures of demographic behavior — mortality, nuptiality and fertility — which he interprets as indicators of the degree of change in sociocultural behavior after crossing the ocean and meeting new agrarian conditions. In Vik, there was little arable land, and the opportunity to own land was limited by traditional property structures. In America, by contrast, the opportunity to own a larger farm with arable land was available until the end of the 19th century and they could thus achieve greater prosperity. Following Malthus' argument, Sunde suggests that these circumstances in North America would lead to a lower mortality rate, lower age at marriage and higher fertility than among relatives and former neighbors who remained in Vik.

Based on a combination of aggregates and microdata, adding especially hard-to-identify runaway sailors, Eide and Thorvaldsen (2011) summed a total of 960,000 Norwegian emigrants to America during the 19th and first half of the 20th century. We might consider this a reasonable compromise between the old and lower emigration statistics and Mørkhagen's (2009) revised estimate of about a million emigrants in his overview of Norwegian emigration (Thorvaldsen, 2018b). Another disputed under-enumeration issue concerns the US census from 1870: a number of cities at the time complained about under counts. Many estimates for under-enumeration lie around 2–3%, except for significantly higher proportions in the southern states. Results for Norwegian immigrants in 1870 based on a combination of aggregates, immigration and emigration lists as well as projections of mortality based on microdata, suggest an under-enumeration of about 10% (Thorvaldsen, 2018b).

Using the linked pairs of censuses from Norway and the USA, a group of economists (Abramitzky, Boustán & Eriksson, 2012, 2013) published research on the motives behind the emigration from Europe to the USA and "the return to migration" — the economists' expression for whether the emigrants' relocation improved their economy. During the period of mass emigration, the United States maintained a relatively open border for immigrants from many countries, making it feasible to study a migration process less hindered by legal entry restrictions. Using linked data on 50,000 Norwegian men, they studied, on the one hand, the effect of wealth on the probability of internal or international migration in the period 1850–1913. Here, they took advantage of variations in the parents' wealth and in the expected inheritance according to birth order, gender composition of siblings and region of residence. They concluded that relative prosperity made the decision to emigrate less likely in this era, suggesting that the poor might be more likely to relocate if migration restrictions were lifted today, and discussed the implications of such historical discoveries for developing countries. Second, the same researchers

estimated the extent to which emigration paid off by comparing emigrants from Norway to the United States with brothers who lived in Norway towards the end of the 19th century. The comparison suggests that "the return to emigration" was relatively low and notes that the relatively poor from urban areas emigrated more frequently. Results from the above mentioned Ullensaker project indicate that the latter result does not apply to Norwegian rural areas, although this can be explained by urban-rural differentials and be due to period effects (Koren, 1979). We therefore need further research with longitudinal data to test the economists' conclusions that relative prosperity among Norwegian residents counteracted emigration and that leaving for the US only paid off to a lesser extent.

### 3.3 IMMIGRATION TO NORWAY

Three volumes written in Norwegian cover immigration history comprehensively together with a summary volume in English (Brochmann & Kjeldstadli, 2008; Kjeldstadli (Ed.), 2003). As part of the first volume, a project analyzed the surnames in the 1801 census as proxies for the missing information about birthplaces in this early nominative enumeration (Sogner & Thorvaldsen, 2001). The surnames originating in Denmark and the German realm mirror the steadily tighter union between Norway and Denmark with the latter as the dominating partner responsible for an influx of administrators and other specialists to Norway. The German names found in the mining towns of south-eastern Norway recruiting miners from central Europe and in Bergen on the west coast with its Hanseatic trade were also expected, while the significant proportion of foreign names, including Swedish ones in the more isolated Northern Norway was more of a surprise. Naturally, many of these non-patronymic surnames had been inherited over the generations and so only indirectly indicated individual immigrants.

Marta Gjernes (2004) explored the immigration of Jews into the Norwegian capital from 1851 onwards when the constitutional ban on their residence in Norway was amended (the 1814 constitution had blocked Jews' access to the realm). The population censuses were main sources, providing information about both religious affiliation and domicile. Norway together with Canada have long traditions of asking about religious affiliation in their censuses (Thorvaldsen, 2014). During the first 30 years after the opening-up, fewer than 100 Jews settled in the capital Kristiania (Oslo). Her thesis, written to support the Immigration to Norway project, studied the 734 Jews settling or born in the capital from 1851 to 1900. Most were businessmen from Denmark and Germany, and many lived in Kristiania temporarily. With the immigration from Eastern Europe beginning around 1880 the number increased, and by the turn of the century, about 500 Jews had immigrated, the majority of whom settled in in the capital. The eastern European Jews came from small villages in rural areas in the north-western part of the Jewish Pale, the region of allowed settlement in Russia, where they had lived as peddlers, shopkeepers or craftsmen. In Kristiania, commodity trade became the economic niche of the Jewish immigrants, and their residences were clustered ethnically. The Danish and German Jews settled in relatively prosperous areas, while the eastern European Jews tended to settle in a less prosperous area near the city centre, where most Jewish social activities took place. Those established helped newcomers with housing and work. The majority of eastern European Jews started as peddlers, but experienced upward social mobility, both individually and between generations. Few of them became wealthy, the majority ending up in the lower middle class, but could afford their children's education. The children started their careers at a higher level than their parents and settled in more prosperous areas. The flourishing Jewish community came to an end when about 800 persons were deported to the German termination camps while other Jews managed to escape to Sweden.

In summary, internal migration has been studied with cross-sectional and linked data on the individual level, particularly around Oslo, inside Middle and Northern Norway and from the southern to the northern regions. Emigration was already well studied based on aggregates, but from the 1970s onwards trans-transatlantic migration from several localities was analyzed with microdata. Our census microdata assisted the writing of the three-volume immigration history and made possible the study of particular groups of immigrants, such as the Jews. A major finding from individual level studies is that integrated parts of whole societies were transplanted due to the social multiplier effect, both domestically and across the Atlantic.

## 4 MORTALITY

### 4.1 SOURCES, RATES AND ENVIRONMENT

Prior to the employment of microdata, Statistics Norway based their studies on historical series of official statistics. In particular they studied mortality, finding that this factor affected population size significantly more than net emigration in most years (Statistics Norway, 1995, table 3.13). During the 18th century, the crude mortality rate fluctuated greatly between 22 and nearly 30 per 1,000. Norwegian historians have disagreed about the relative role of poor harvests and epidemics. In the first half of the 19th century, the variations of the general mortality were less striking, with a crude mortality rate of 18.9 per 1,000 shortly after the Napoleonic Wars, and 17.3 per 1,000 in 1851–1855. During the following hundred years, the general picture has been one of a consistent downward trend although with frequent short-term fluctuations, especially before 1900. About 1900 the crude mortality rate was 14.5 and fifty years later 8.5 per 1,000, a reduction of 41% (Backer, 1961).

Official infant mortality rates for Norway based on ecclesiastical reports have existed from 1836, but until the 1850s only at the national level. Based on statistical techniques it has been possible to project rates back to 1735, deriving from available demographic aggregates of births and deaths (McCaa, 1989). Around 1800 an infant mortality rate of around 180 per 1,000 live births is estimated, which by the mid-19th century fell to around 100 per 1,000. As part of the rapid decline in the crude mortality rate witnessed after 1900, infant mortality fell sharply, from around 100 to 22 per 1,000 in 1950 (Backer, 1961). The geographical variation in infant mortality shows a coastal inland pattern, with higher mortality along the coast, as well as a higher mortality in the north, compared with the south (Sommerseth, 2003; Thorvaldsen, 2002). A marked difference between urban and rural infant mortality has also been unmasked (Thorvaldsen, 2002). Official statistics show that the urban mortality rate remained 25–30% higher than the rural infant mortality throughout the period 1856–1920, both declining slowly at first, and sharply after 1900. From this time the urban and rural infant mortality rate series converged. However, overall urban mortality declined relative to the rural rates from 1920. A central question has been to investigate who paid for the urban penalty and who received the rural reward.

Norway is located in the western and northern part of the Scandinavian Peninsula, and its lengthy topography divides the country into different climatic zones, extending from warm to cold temperate zones, and polar climate in the mountains and the northern regions. In addition, the inland regions have a more continental climate while the coast have milder winters and cooler summers. An illustrative example may be the maximum monthly average temperature of 22.7 °C (measured in Oslo in July 1901), and the lowest monthly average temperature of -27.1 °C (Karasjok in February 1966). Being dependent on the harvest from the soil, fjords and ocean was a yearly risk, and once winter arrived, it affected draughty houses, and the frost lay like a claw until spring arrived. Under such circumstances, the immune system weakens, and the risk of infections increases.

Already from 1820 onwards, the church burial registers included an option to report infectious diseases and accidents specifically. In 1877, it stated that all causes of death had to be registered including information about whether the medical doctor had visited the deceased prior or upon death. The registration of causes of death was part of the medical statistics, provided not only by a priest, but also, from 1853 onward, increasingly by medical professionals. However, around 1860 only about 40% of all deaths in Norway were represented in the medical profession's reports, increasing to around 80% in 1900 (Pedersen, 2007). The authorities obviously had an assistant in the local priest, and this gives us a unique insight into long-term trends in causes of death on the individual level.

Well into the 20th century, infectious diseases were significantly more prevalent compared to other cause-of-death groups, and the above-mentioned urban-rural mortality difference has been explained by the fact that the scattered settlement in rural areas has been considered a protective function against the spread of diseases. Conversely, the population density in the cities was accompanied by a greater risk of adverse exposures. Already in the mid-18th century, towns started to monitor the public health, and preventive measures were implemented. For example, the city of Trondheim got the country's first public waterworks in 1777. However, a study based on the burial records for one of its parishes shows that the effect on the crude mortality rate was marginal (Knudtsen, 1997). A more immediate effect, however, was observed with the abolition of the eight days deadline for church baptism, carried out in 1771, allowing parents to have their new-born baptized at home, and not in draughty churches during the winter season (Knudtsen, 1997).

## 4.2 INFANT MORTALITY

Over the course of the 19th century, the risk was highest in winter for those infants who had survived their first six months of life (Fure, 2002; Sommerseth, 2003). The consequences of generally poor sanitary, hygienic and living conditions are important factors when trying to understand the higher risk the winter season had for the survival chances of the infants.

In addition, the typically short and hectic length of the harvest season could potentially create a sequence of severe risks. Short harvest seasons required all household members to participate, and this was likely a competing activity for women who breastfed new born children. In a study that covers a southwestern parish named Etne and utilizing linked nominative records, it is argued that the peaks in infant mortality are best explained by farmers' wives occasionally reducing breastfeeding due to their participation in seasonal agriculture labour (Dyrvik, 1997). For the coastal women, the amount of work increased with her husband's absence during the large seasonal fisheries, and at times the workload must have been difficult to reconcile with breastfeeding, potentially creating an irregular nutritional practice (Sommerseth, 2006; Thorvaldsen, 2002). Nutritional practice could also be gendered, as shown in the Trondheim case where infant mortality rates from the late 18th century show an unusual excess mortality among boys compared to girls during the neonatal period, which is explained by a culture of favouring the infant boys with a heavy farmer's diet instead of breastmilk (Knudtsen, 1997). In a summary research article about infant mortality in the five Nordic countries, the level and consistency of breastfeeding as well as child care more generally were stressed as decisive factors behind mortality patterns transcending the national borders (Edvinsson, Garðarsdóttir, & Thorvaldsen, 2008).

During the 18th century, Norway witnessed a social cleavage, when the cotters appeared alongside the dominating social group of farmers. During the 19th century, however, a social levelling-process happened, partly driven by more available land due to mass emigration to the US, but also due to increased demands from a commercialized market which gave the possibility for supplementary income. In Rendalen, a parish located in the southeast of Norway, an extended database version of Sølvi Sogner's family reconstitution was used for several studies. The infant mortality shows the "expected" social gradient, with higher mortality among infants of the lower social class compared to the upper class. During the 19th century, however, these rates converged, probably mirroring the increased importance of timber trade in that area (Sogner, 2000). In Etne, a different picture became apparent, with higher mortality among infants of farmers compared to cotters, but similarly as in Rendalen, the gradient disappeared during the 19th century (Dyrvik, 1997). A study of infant mortality in Asker and Bærum (southwest of Oslo) revealed no significant differences according to social class (Fure, 2000). Clearly, these divergent patterns confirm the complex nature of the causal forces affecting mortality risk.

## 4.3 OVERALL MORTALITY AND CAUSES OF DEATH

If we include all age groups, the mortality patterns do not become clearer. Contradicting McKeown's famous thesis on the relationship between nutrition and mortality, a study based on church records from 44 parishes linked to the 1801 census shows that Norway in the early 1800s had higher mortality among the upper social class compared to the lower social strata (Engelsen, 1983). A possible explanation for this can be found in the socio-geographical pattern typical for parts of Norway, where the poorest often lived on a small cotter's place in the periphery of the main farm, and the wealthiest lived more centrally in so-called "klyngetun" [cluster yards — not to be understood as villages], where the farm's houses, as the concept suggests, stood close together. Consequently, the exposure to infectious diseases was higher in the social upper class and to such an extent that it resulted in their higher mortality (Engelsen, 1983). Other studies have, on the other hand, emphasized that the center-periphery (or urban-rural) dichotomy may miss the dynamic and seasonal shift in settlement patterns. This is often found along the coast where seasonal migration was enormous during the big seasonal fisheries, and thus could potentially create regular surges of infectious diseases for a short period in typical rural areas (Sommerseth, 2003).

Cold and wet autumn and winter months were also a risk for the whole population. The disease environment in Trondheim city during the last part of the 19th century was dominated by airborne infectious diseases, with pneumonia and bronchitis as two of the largest cause-of-death groups. Airborne diseases killed nearly 60% of those children who died, and was the main killer for ages 1 to 49 in a study of 19th century Trondheim burial records (Sommerseth & Walhout, 2019). Poor and crowded housing, deficient sanitary and hygienic conditions, also contributed to high tuberculosis (TB)

mortality. Placed in the national context of relatively high TB mortality, a recent study based on church records takes a closer look at age and sex differences, as well as societal and spatial TB inequalities in Tromsø during the years 1875 to 1920. The results, though not statistically significant (probably due to small numbers), show the expected female excess mortality, and women aged 15–49 were the most exposed. Interestingly, the study also finds higher mortality among the working class, and in areas where dwelling quality and overcrowding was an issue, women were particularly vulnerable. One explanation for this pattern could be that women (and children) spent more time inside these dwellings than men who spent long working days elsewhere (Kovacevic, 2020).

Traditionally, the 1918 flu pandemic was perceived as socially neutral. However, Mamelund (2006) showed how it affected the lower social classes more severely, leading to higher death rates. His results are based on burial records linked on the individual level to annual city censuses. By comparing two socially contrasting parishes in Kristiana (Oslo), namely Frogner and Grønland-Wexels, he found that apartment size (which was an ideal proxy for rent ( $r = 0.98$ ) and therefore deemed a good proxy for income), was negatively and significantly associated with mortality. All other factors being the same, his results showed a 49% higher mortality rate for those residing in the poor parish of Grønland-Wexels compared with the affluent parish of Frogner.

#### 4.4 MORTALITY AND ETHNICITY

In an ethnic context, studies show higher infant mortality among Sami nomads compared to other ethnic groups, and one explanation for this may be the strains of moving between summer and winter grazing (Andresen, A. 2001). This explanation is also in line with studies of infant mortality in Jokkmokk (Sweden) where the high infant mortality among Sami nomads coincided with the most intense work period on summer pasture (Brändström, 1988). Based on linked church records, results from the sea Sami area of Tana in Finnmark did, on the other hand, not find a significant mortality difference between infants born by Sami compared to those born by Norwegians. This strengthens the explanation that the high infant mortality found among the Sami nomads, was likely a consequence of neglected care due to periodically severe working conditions for women who just recently had given birth, and not driven by ethnicity as a strict biological measure (Sommerseth, 2003, 2006). The same study found higher infant mortality among the Kvens (Finnish immigrants). Studies that cover the historical population of northern Finland explain the high infant mortality (20–40%) with artificial nutrition — usually fresh or sour milk (Brändström, 1988; Lithell, 1988; Pitkänen, 1983). An obvious assumption is therefore that some Kvens brought with them a cultural nutritional practice when they migrated to Norway (Sommerseth, 2006). The Kven migration to Tana in Finnmark took place mainly from the middle of the 19th century, and we can clearly see how the trend over time shows a declining mortality from a relatively high level during the first years of settlement, to a level similar to the Sami and Norwegians one or two generations later.

According to Sunde's family reconstitution (2001), the first generation of immigrants to the US Mid-West were faced with a process of acculturation, and this in combination with a more fragile social and kin network, different nature and climate, family households met new challenges, some for better and others for worse. During the settlement process children were born, and it is reasonable to assume that infants of the first-generation immigrants were exposed to a higher risk just by being born during a process of securing permanent footing in a new country. This assumption is supported by studies done on Norwegian immigrants in the USA, where the first generation of immigrants experienced a higher risk of losing a child before the age of one.

#### 4.5 EXPLAINING THE INFANT MORTALITY DIFFERENTIALS

In Norwegian historiography, there are various explanations for the relatively high 18th and early 19th century infant mortality, its comparatively low level in a European context, and its trend over time, which was characterized by a steady decline in the 19th century and a sharp decline in the 20th century. There is a broad agreement that public health measures in the 19th century through vaccination programs, a focus on stillbirths, introduction of educating midwives in 1810, appointment of district physicians and establishment of health commissions (an urban phenomenon from 1803, but fully incorporated in all municipalities of the country from the 1860s), were all conditions that stimulated a low mortality level (Sogner, 2000). On the other hand, what caused the initial fall in infant mortality has been questioned. For example, the work of doctors and midwives may or may not have been pivotal for the early phase of the decline in infant mortality, since the decline began long before

the districts had their own doctor and midwife (Dyrvik, 1997; Fure, 2002). As demonstrated in a study of infant mortality in Hammerfest town utilizing the ministerial records, it is likely that education and presence of a health care system resulted in an earlier and more beneficial effect in urban compared to rural areas where in the latter case a couple of doctors and a midwife had to cover vast distances (Mølmann, 2004).

The ethnic dimension is also important when we examine the potentially beneficial effects of health care measures. With increased forces from the late 19th century, Norway executed an offensive policy towards the Sami population and later the Kven population in the so-called Norwegianization process that sought complete assimilation. The northernmost and ethnically mixed county of Norway, experienced one of the highest infant mortality levels compared to other counties and is today faced with the highest crude death rate. Questions remain about the long-term trend and potentially continuing health effect of the Norwegianization process. One explanation is the long-term effects of sub-standard living conditions during childhood and adolescence (Forsdahl, 1973).

In recent decades, studies concerning the reproduction of health advantages and disadvantages across generations have flourished. One study that prepared the ground for further research in Norway, showed an important turning-point with improved life expectancy among women already from the 1770s, while it took more than 80 years before men started to catch up with women (Mamelund, 2001). Again employing her Rendalen database, Sogner (2000) who already then was acquainted with Mamelund's results, argued that the improvements in women's health and life expectancy had a beneficial effect on their infants. She further argued that the difference in life expectancy between men and women was a result of the introduction of commercialized forestry on a large scale from the 1790s. This meant rougher living conditions especially for men. However, women, who in general were not employed in forestry, may have profited indirectly from the new sources of family income, beneficial to a healthier lifestyle and thereby improved survival of their infants.

To what extent a mother's health matters has also been questioned by looking at the living conditions when a mother was born and the possible effects on the survival of her children. In other words, were mothers born in difficult years imprinted by the adverse conditions so that when giving birth, their infants had an increased risk of dying? Good and bad periods were defined by using levels of high and low infant mortality, and the results showed that the neonatal mortality was twice as high for infants whose mothers were born in years with high infant mortality. After the neonatal period, there were no mortality differences according to mothers' birth year. A possible explanation is that some of these mothers themselves were born under adverse conditions, caused by disease or undernourishment, in utero or in early infancy. They might have been predisposed to bear weaker infants (Fure, 2002). This transgenerational programming may have happened epigenetically, as changes in the gene expression system that can be inherited from mother to child (Bygren et al., 2014). Along with increased access to longitudinal data that covers multiple generations, we are now in a position where these questions can be given robust answers, also in an international and comparative fashion.

The international Intermediate Data Structure (IDS) format was applied in a comparative study of intergenerational inequality in infant mortality, in which research teams from the Netherlands, Belgium, Sweden and Norway participated with their respective datasets. In the first part of the project, separate analyses were carried out, but with a common theoretical and methodological design. The results from linked ministerial and census records in the Historical Population Register for Troms province in Norway, show that there was a transmission between the generations with respect to risk of experiencing infant deaths. A woman's children were at greater risk of dying before the first birthday if their grandmother had had several infant deaths. The risk of infant death among children of daughters from such high-risk families was at least 30% higher than among infants born to the daughters of mothers who had experienced zero dead infants. At the opposite end, we find the majority, where 60–70% of families across generations, never experienced the loss of any infant (Sommerseth, 2018). Similar findings were also made in the analyses of the other countries, and overall, these findings point in the direction of a different understanding of how mortality risks varied among the very young (Quaranta & Sommerseth, 2018). Infant mortality in historical populations did not affect every house and family. The next step in the project has been to assemble all the countries' datasets into a common database and carry out a common statistical analysis. The results confirm findings from the first round and represent a pioneering work for further internationalization and comparative historical demography (Quaranta et al., 2017).

In sum, Norwegian mortality studies have focused on rural-urban differences and infant mortality, using the rich collection of transcribed vital records from the church books. Recent historiography especially dealt with long-term influences over three generations from grandmothers to grandchildren and how trauma in the mothers' early life could influence the survivability of their small children. Other studies deal with the situation of the ethnic minorities and the migrants.

## 5 NUPTIALITY

The nuptiality studies have focused on the timing of marriage relative to the transfer of property between the generations, illegitimacy and intermarriage between the ethnic groups.

### 5.1 MARRIAGE MARKETS AND TIMING OF MARRIAGE

The effect of shifting generations and mortality levels was the topic of Hans Henrik Bull's dissertation on nuptiality, presenting a long-term study of social homogamy, timing of marriage and illegitimacy. This is yet another project researching rural Rendalen close to the Swedish border in the 18th and 19th centuries (Bull, 2005, 2006) building on Sølvi Sogner's database with longitudinal information (Gjelseth, 2000). Prior to 1870, the occupation of parents was to a varying degree recorded in the Norwegian parish registers, but can usually be found in the censuses. Structural changes over time, which led to an increase in the number of farm workers, reduced the degree of homogamy among farmers, and extended the marriage market for farm workers, thereby increasing homogamy among the latter group. Controlling for these structural changes, it is clear that social boundaries between farmers and farm workers prevailed at least until the end of the 19th century. Multivariate analysis identified family characteristics that led young men and women to marry homogeneously. The farmers, especially, exerted influence on their eldest sons to marry farmers' daughters, but the role of the father in the mating process also secured economically viable partners for the other siblings.

Furthermore, Bull's dissertation analysed the connection between inheritance and the timing of marriage as well as the development of illegitimacy rates in the same rural stem-family community. In pre-industrial Europe, access to a livelihood, whether on the family farm or other means of earning a living, was considered a prerequisite to marriage, usually through intergenerational transfer, which, especially in the 18th century was affected by the level of adult mortality. Early mortality among adults created opportunities to marry a widow(er), and for a farmer's eldest son to take over the farm. In the 19th century, a reduction in adult mortality and the subsequent rate of remarriage closed this opportunity. However, the timing of marriage was less affected by the intergenerational transfer of wealth than by the family's ability to use the work force of men and women. Thus, the amount of wealth a woman brought into the marriage might be valued less than her ability to work and run a household.

In the second half of the 18th century, the number of illegitimate births increased in Europe, while the number fell a century later. Bull's dissertation examines premarital sexuality and marriage in the Rendalen database between 1750 and 1900. A new measure of premarital sexuality, called the extramarital pregnancy rate, includes children born out of wedlock as well as children born to pregnant brides. The study shows that both among the daughters of farmers and of farm workers there was a clear increase in extramaritally conceived children at the end of the 18th century. This was caused by a development whereby the choice of partner was now made based on mutual attraction rather than economic strategies. The fall in illegitimacy at the end of the 19th century was likely connected to the bourgeoisie family's more puritan view of female sexuality spreading among the farmers, thus distancing themselves from the group of farm workers, and once more strengthening their control over the children's marriages in the second half of the 19th century (Bull, 2005, 2006).

### 5.2 NUPTIALITY AND ETHNICITY

Ethnic marriage patterns were studied in two parishes north of Tromsø from 1770 to 1900 (Larsen, 2008). Microdata and aggregates from the church records and censuses were used to define ethnically homogeneous and mixed areas, in order to compare their demographic developments with individual level data. From 1845, the Norwegian censuses contain ethnicity information, usually denoted as

nationality, primarily to enumerate the Sami and Finnish minorities. When the censuses became nominative in 1865, ancestry and belonging to an ethnic group were the main criteria, but towards the end of the century, individual and cultural criteria like language became more central. The interpretation of the ethnic markers are not always straightforward, for instance the ethnonym "Fin" can mean Sami in some censuses and Finnish in others, but as a rule close inspection of the nominative manuscripts reveal the correct interpretation (Thorvaldsen, 2018a, p. 138).

The phrase "meeting-place of the three tribes" has been coined to describe the ethnic composition of the two northernmost provinces in Norway. This is a reference to the mixed ethnicity of the region with people of Sami, Norwegian and Finnish heritage. While it is contested which areas were traditionally Sami, however, there is no doubt that Norwegians took over large areas formerly used by the Sami. Thus, the Norwegians spread from the south towards the north and from the coast towards the inland. The Finnish immigration from Sweden and Finland started in the 18th century but gained real momentum from the 1830's. The 1865 and 1875 censuses show that 13% of the population in Troms province was Sami, while 5–7% were of Finnish stock, proportions decreasing to 9% (Sami) and 2% (Finnish) in the 1900 census. Concurrently, the ethnically mixed population grew from 3% to 9%. Research confirms that the censuses render ancestral ethnic heritage quite correctly, so these figures mirror the intermarriage of the ethnic groups. Mixed marriage between the Norwegian and the Sami group was relatively infrequent, while the Finns married both Sami and Norwegian partners, functioning as a catalyst in the ethnic salad bowl. However, even if individual members of the three ethnic groups could be on good terms, there are numerous reports about conflicts, especially between Sami nomads and Norwegian farmers.

As expected, almost all women in the two parishes north of Tromsø married and most of them very young around 1800, as they did in other Sami-dominated areas. This pattern disappeared early in the 19th century, perhaps due to economic pressure. Geographical boundaries increasingly constituted ethnic boundaries, and marriage contact between the parishes became infrequent, especially between the Sami population groups and particularly after 1880. The demographic developments in the parishes were different, with respect to marriage age, proportion unmarried, remarriage frequency and population stagnation or expansion. A stricter ethnic hierarchy, discriminating the Sami, caused gender imbalance and thus lowered the marriage rate in one of the parishes. This is the likely cause of increased illegitimacy outside of the area with strict Laestadian religiosity. The Laestadians, while still members of the State Church, attempted to regulate all aspects of their adherents' lives. Laestadianism also counteracted some of the discrimination of the Sami. In the population exposed to Norwegianisation, Sami ethnicity tended to dominate and influence the spouse. This may be because the ethnically mixed population here had no choice but to stick to the Sami identity (Larsen, 2008).

Nuptiality has been analyzed in selected localities. To sum up, in the south-east, marriage was connected to the transfer of property and there was a development from arranged marriages to self-selected partners in the early 19th century, concurrent with a rise in the number of illegitimate births. In another highlight from the north, "the meeting place of three tribes", there were more restrictions on marriage between the Sami and the Norwegians than on the Sami marrying immigrants from Finland. The two latter groups were able to combine access to land with agricultural experience, while the two former groups often had conflicts about using land for traditional reindeer pasture or agriculture.

## 6 FERTILITY

### 6.1 PATTERNS OF DECLINE

Influenced by the Princeton project (Coale, Anderson, & Härm, 1979), Sogner, Fure and Randsborg (1984) mapped the Norwegian fertility patterns both on the regional and municipal levels. Based on official statistics, their results point to a fertility increase in the period 1801–1835, a stable high fertility in the period 1835–1900 and a fertility decline starting around 1890. Moreover, this nationwide study suggests that the spread of birth control had both a social and geographical gradient, starting among couples in the higher social classes in urban environments (Sogner et al., 1984). In the following we shall refer to studies that used family reconstitution primarily based on church books and censuses, where fertility behavior within the boundaries of a parish or municipality have provided us with a more nuanced picture of the diffusion theory proposed above.

Generally, both marriage and fertility — two closely related phenomena in historical populations with limited access to contraception — were positively correlated with social status and wealth. With easy access to a livelihood, the earlier the marriage, the more children born. Consequently, for the rural areas in Norway, peasant women had a lower age at marriage and higher age-specific fertility compared to the cottier women (Sogner, 1979). In the pre-transitional phase, accompanied by an overall rapid population growth, Norwegian society underwent radical economic and social changes, with a strong development in the secondary and tertiary economic sectors both in rural and urban areas, growth in the number of cities and suburban areas, increased internal migration, all pointing to a complex web of driving forces behind the fertility transition phase. Undoubtedly, major societal changes affected demographic behavior, and as have been demonstrated in several master theses conducted on the parish level, the abovementioned social and geographic gradient have numerous exceptions. For instance, there was a relatively higher age at marriage and low fertility among peasant women in Norddal and Rendalen (Linge, 1977; Sogner, 1979); equal fertility levels between peasants and cottiers in Ullensaker and Bø in the region of Telemark (Halvorsen & Indseth, 1875; Randsborg, 1979), while the decline started first among the lower social classes (cottiers and farm laborers) in Kolbu (Fure, 1980). A positive correlation between fertility and social status was on the other hand found in the cities Christiania (Oslo), Bergen and Trondheim. "Own children" rates calculated from the 1801 and 1875 censuses indicate that fertility was somewhat higher in the upper than in the middle and lower social strata. For the latter two groups, fertility was virtually identical (Sogner et al., 1984)

A contradicting result was found in the city of Stavanger. Based on a random sample of 609 families drawn from the censuses of 1920 and 1930 linked to demographic data from church books and the city population register, Dyrvik and Alsvik (1987) studied the introduction of birth control during the first decades of the 20th century. Their analysis showed no connection between birth control and occupation, income or standard of housing. Fertility decline started in the humble quarters of Stavanger, but accelerated in more affluent sections. The adoption of fertility regulation differentiated most notably between the trades, the living quarters and the parental birthplaces. This indicates that the innovation spread through a hierarchy of communities, using a communication network defined by cultures and subcultures.

## 6.2 ILLEGITIMACY

So far, we have focused on the fertility behavior within marital unions. Left out are the illegitimate births. Around the beginning of the 19th century, regional numbers show that between 2 and 10 out of 100 births were registered as illegitimate, and our narrative would be deficient if we excluded them. Based on church books from Moss town covering the years 1776–1825, Eliassen found an increase in illegitimate births from 10% in 1780s to 15% during the first decades of the 19th century (Eliassen, 1981). Comparing the increase with the concurrently declining proportion of marriages, two distinct marriage strategies were detected. During the first period, until 1800, marriages were postponed. During the first decades of the 19th century, however, illegitimate births to a lesser extent led to marriage. It also varied to what degree different priests let the prospect of marriage influence notions about legitimacy.

In her master thesis, Haavet (1982) examined the social status of unmarried parents before and after an illegitimate birth. Her study is based on 44 parish record books covering the years 1795–1800 and 1802–1803, linked to social status in the 1801 census (cf. section 2.1). Relative to different social classes in the unmarried population, there were slightly more servants who gave birth to illegitimate children. Unmarried parents who were registered as living as single in their childhood home, had on average occupations with a lower social status than the general population. This was especially true for the mothers, while the distribution for the fathers was more even. The thesis concluded that there was little evidence pointing in the direction that unmarried parents were punished for their premarital intercourse, for example by being defined as outcasts after an illegitimate birth. On the other hand, if the illegitimate birth did not lead to marriage (although it often did), the chances of a later marriage to someone else were highly skewed by gender. Perhaps not surprising, unmarried fathers later on had a marriage frequency that seems unaffected by an illegitimate child, while single mothers were handicapped in the marriage market.

The fertility decline, in summary, started in the higher urban social classes and was slowest in the peripheral, agricultural environments. Results from local reconstitution studies are, however, more nuanced, indicating that the fertility transition was a cultural phenomenon which spread through a

hierarchy of communities. Illegitimacy rates varied throughout the country, a result of mating cultures that could be tolerated locally to varying degrees, and priests who often modified their classification of illegitimacy if they were positive that the parents intended to marry.

## 7 FAMILY HISTORY

### 7.1 INTERGENERATIONAL CO-RESIDENCE

The prevailing discourse on family history since the early 1970s has downplayed earlier assumptions that industrialization and urbanization diminished intergenerational co-residence. Accordingly, the dominant form of household structure in Western Europe and the United States in pre-industrial times was nuclear, and industrialization did not change the structure of domestic life. A typical life cycle pattern is one where elderly parents gradually moved into one of their children's households when they were no longer able to keep an independent household. Norwegian scholars have largely confirmed the dominance of nuclear households in pre-industrial times, with studies emphasizing the economic role of the household as a unit of production and consumption around the nucleus of the 18th and 19th century farmers, cotters, and fishermen households (Bull, 2000; Dyrvik, 1993; Fure, 1986; Sogner, 1978, 1990; Solli, 1995). Moreover, the majority of these studies found that occupations connected to owning property (mostly farmers) were associated with an increase in stem-families compared to landless cotters, and thus became a focal point in studies on the connection between land ownership and extended family household formations in the Norwegian pre-industrial context. The allodium law and residing (*åsetes*) right played a crucial role in maintaining a stem family system in Norway, however strongly associated with a sustenance economy.

In addition to stating that the oldest son should inherit the farm (primogeniture), allodium law described a set of rights that the older generation had upon retirement and the transfer of the farm. After moving out from the main building into a separate house ("kårhus"), the older generation could claim a part of the farm's livelihood to provide security during their old age. Thus, among family historians, the allodium law and the retirement contract accompanying it have been seen as strong evidence for a consistent and culturally homogeneous family system across pre-industrial Norway. Sogner suggested that the strong influence that laws of inheritance and land transfer had on people's choice of living arrangement met its end during the 19th century, when competitive alternatives arose. New possibilities were migration to America and Northern Norway, and to new occupation alternatives in the towns (Sogner, 2009).

### 7.2 FRAGMENTATION OF THE HOUSEHOLD STRUCTURE

In his dissertation on the history of the household, Solli (2004) investigated three related research questions: To what extent did household and life cycle patterns change during the 19th century? In what way did this change, and why? The empirical and methodological focus was on the life course leading up to the establishment of households and the defining marriage, and the sources were the full-count, nominative 1801, 1865 and 1900 censuses used cross-sectionally. The initial analysis on the national level identified typical life cycles, household characteristics and a range of social, political, economic, cultural and demographic explanatory variables. These explanatory variables were tested against the life cycle and household patterns, suggesting three sub periods: until ca. 1840, 1840–1870 and 1870–1920. The first phase was characterized by a stable supply of land resources being the *sine qua non* for household establishment, such as the northwestern European household system outlined by Hajnal (1965). However, wars and crises together with increased demand for Norwegian exports led to a somewhat higher number of varied livelihoods. Major changes in the resource base, for example more extensive herring fisheries, facilitated the splitting of some farms into peasants and cotters' smallholdings.

The second phase, ca. 1840–1870, was more adaptive, with marriage becoming less contingent on a stable livelihood. The pressure on the household structure was both demographic and economic with population growth and expansion of trade and commodity production. Along the coast, the population grew rapidly, based on shipping, trading and fishing. However, the growth did not lead to major changes in the household structure; people settled traditionally by increasing the number of smallholdings and families resulting in an average of five to six persons. However, the adaptation

was different where cattle and grain production dominated. In mountainous parts, the population increased less, in some areas it even declined or stagnated. Siblings and the retired ("kårfolk") more often established their own households, and the former established themselves with the few new types of livelihoods such as teachers, shopkeepers, day-workers, with incomes often too small to feed an entire family. Thus, in the uplands, the traditional household structure was destroyed during this period. In spite of industrial growth, the urban changes in household structure were confusingly similar to the changes in mountainous Norway. A starting fragmentation of the household structure characterized both, with more single persons in all age groups, especially the young (20–29) and old (over 60). Also, more women of 45 years and older became household heads, resulting in more production-free livelihoods that were less dependent on a married couple. Instead, working in the expanding trades after the economic liberalization of the 1840s became an increased opportunity. Changes in the rural flatland communities in eastern and mid Norway were less significant. Male servants still contributed heavily to the work force and were often part of the farm household. Thus, rationalization and specialization in agriculture influenced life cycle and household structures relatively little, maintaining the sharp social differences.

In the third phase (ca. 1870–1900), the relocation of social production out of the household to the more heterogeneous secondary and tertiary sectors became extensive, changing the life cycle and household structure significantly in most of the country. Young single men were main players in the waged labor market, while women were main characters in the households. The larger the secondary sector, the lower the age at marriage and establishment, the longer the children were raised at home and in schools, the smaller the household size and the more people were single. Thus, the third period was characterized by a quantitative socio-economic and demographic shift with more production-less households established in densely populated areas (Solli, 2004). With hindsight, we must conclude that Hajnal's theory of typical and interrelated household markers, fit the Norwegian reality better at the start than at the end of the period studied (Szołtysek, Ogórek, & Gruber, 2021).

The inheritance practice described above assumes a life change for the older generation. Sommerseth (2011a) has explored family living arrangements from the perspective of the elderly and their co-residing behaviour with an own adult child in Northern Norway during the last part of the 19th century based on the 1865, 1875 and 1900 censuses. As stated by Ruggles, this perspective has an additional advantage: it controls for demographic conditions and constraints typically found in historical populations of the western part of the world. Late age at marriage, combined with no deliberate fertility limitation, resulted in generations tending to be long (Ruggles, 1994). In addition, the combination of long generations and relatively short life expectancy reduced the potential number of years it was theoretically possible for adult children to live with their parents. Furthermore, since married brothers and sisters seldom resided together, although the relatively high fertility resulted in family living arrangements that extended beyond the nuclear form, these would always be in the minority.

### 7.3 ETHNICITY AND DECLINE IN INTERGENERATIONAL CO-RESIDENCE

A case study of Northern Norway was also needed to test the narrative of a culturally homogeneous Norwegian stem family system. Pre-industrial Norway was not culturally homogeneous. Northern Norway's population consisted, and still consists of three different ethnic groups: Sami, Norwegian, and Kvens. Contrary to primogeniture, an alternative common cultural practice existed among the Nomadic Sami as well as among the Coastal Sami population, an ultimogeniture inheritance practice. This means preference was given to the youngest child — preferably to a boy — when property was transferred between generations. The youngest child then provided for his or her parents throughout their lives. The presence of two different inheritance practices within the same geographical area gave a unique opportunity to compare the effect such practices had on living arrangements for a population that shared similar ecological and economic opportunities.

It is well known that in a pre-industrial economic system, intergenerational living arrangements were beneficial both to the older and to the younger generation. Among farmers and fishing farmers, two adult generations in the household meant that capacity for labour-intensive work was ensured. Conversely, in fishing households, the working crew was often organized across the household boundaries. Young sons went out early to participate in the fisheries, and this gave independence in the form of having their own income. In addition to this, the ocean was part of the commons; one could not inherit access to fish there.

By the close of the 19th century, less than half of all elderly people (60+) resided with an own adult child (18+) compared with approximately 60–65% 35 years earlier. Intergenerational co-residence was positively associated with being a married Sami male with an occupation in farming or combined fishing and farming (Sommerseth, 2011a). It is argued that ethnicity played a role; however, its effect disappeared after controlling for economic activity.

The change in intergenerational co-residence was primarily characterized by a decline in the number of married sons staying in their parental homes, and the decline was persistent in all economic sectors. What we see is an increase in widowed and dependent elderly people living as lodgers in households of supposedly non-relatives. Just as the inheritance practice may have expressed an ethnic practice in 1865 and 1875, its articulation was less visible in 1900 (Sommerseth, 2011b). It is further argued that the decline in intergenerational co-residence between 1875 and 1900 may be explained by young men's increased opportunities to take on different occupations, and for some this may have been more attractive than staying at home. Irrespective of the father's marital status, the majority of children living with their elderly fathers were sons. Secondly, we know that the Norwegian state increased its efforts to assimilate the Sami population into Norwegian law and culture. Along with this assimilation process, we also see an increased interaction in the private sphere, for example inter-ethnic marriages. Thus, as ethnic differences in intergenerational co-residence had disappeared in 1900, the consequence was an approach towards a similar family system for the Sami and the Norwegian population.

For one single multi-ethnic parish in Northern Norway census data formed the basis for a longitudinal approach to investigate further son preference. Based on father's complete fertility history, and controlled for any child loss, the study asked if a son preference merely was a consequence of inheritance practice, such as co-residing with first-born son. Although the sample was small, few indications supported this. Again, sons were preferred. The analysis suggests that one important reason is the masculine character of the fisheries, where the transfer of knowledge from father to son was crucial for successful economic results, thus promoting masculine obligations across generations, strong enough to have a significant effect upon living arrangements, irrespectively of ethnic affiliation (Sommerseth & Thorvaldsen, 2016).

As a generalization, these studies confirm the dominance of nuclear households even in pre-industrial times, emphasizing the economic role of the household. In addition to prescribing male primogeniture, the allodium law prescribed rights to housing and subsistence for the older generation upon retirement. A gradual development from the 18th to the 20th century was first characterized by slow population growth and a stable supply of land resources being a necessary condition for household establishment. Exports and rich fisheries in the 19th century with accelerating population growth enabled the establishment of new households based on livelihoods outside agriculture. Next, the period from ca 1870 was characterized by a larger socio-economic and demographic shift with more production-less and smaller households dominated by women in densely populated areas. In Northern Norway, there was a tendency for single fathers to live with a son, facilitating the transfer of occupational skills over the generations.

## 8 A NATIONAL HISTORICAL POPULATION REGISTER (HPR)

The already mentioned grant from the Norwegian Research Council (see section 2.2) accelerated the creation of the Historical Population Register of Norway, covering the whole period from 1800 onwards and connecting with the modern population register, which was introduced in 1964. The historical population register combines microdata into an integrated database with two main objectives. It documents, formats and provides historical personal records covering an increasingly large part of the country at the disposal of researchers, ultimately expanding it to a national and longitudinal register for the last couple of centuries. This interconnection of different sources at the individual level provides new sources for critical insights (Thorvaldsen, 2011).

The HPR places Norway at a vantage point in terms of research on a wide range of questions about the historical development of the population which forms the background for contemporary conditions. Due to the lack of readily available individual level data, in particular the mid-20th century is an understudied period in our population history despite the crucial demographic changes, such as declining fertility and mortality and the post-war baby boom. Other examples with a wider timeframe include changes in family structure, changing migration flows domestically and across the borders and

a revised personal name culture. The possibility of linking such a historical register against the current Central Population Register opens unique opportunities for research on contemporary demographic phenomena with a historical perspective. This is particularly important when inter-generational processes are central, as with social and regional mobility, education and career choices. For researchers in the field of medicine, psychiatry and public health, the possibility of following family relationships over many generations will be a valuable source for studying the heredity of diseases and disorders. As outlined above, Sommerseth has used the HPR to demonstrate transmission of levels of infant mortality between generations.

For legal and ethical reasons, see section 2.2, the HPR must be in two parts. Until the 1920s, most of the information is available to everyone via the Internet and thus a popular offer to local historians, genealogists and locally oriented teaching. Information from the period after the 1920s will be reserved for researchers upon application and ethical vetting. The full count censuses up to 1910 are already made available by means of data processing, and the full count 1920 census will be public on the Internet by early 2022. It is, however, uncertain to what degree we can reconstruct the numerical censuses from 1815 to 1855 at the individual level with nominative data from the HPR. The first version of the 1950 census has been transcribed with machine learning techniques and an improved version will be available for selected researchers in 2022, while work on the 1930 census has started. The baptismal lists can be used until 1930, and data on burials and marriages are in principle open until today, and these church records have been transcribed until the mid-20th century. Statistics Norway has already digitized the censuses since 1960. The censuses since 1910 contain birthdates, facilitating record linkage with data from other sources. Many baptismal, wedding and confirmation lists from the 1800s contain birthdates, and when we shall have certified their quality, they will be linked to the 20th century censuses. Much of this work is automated, because we have access to lists of standardized names; and in addition to this, genealogists contribute with links through the website Histreg (<https://histreg.no>). The census aggregates can also be used to assess the representativeness of researchers' extracts from the Population Register.

We cannot satisfy all potential users of the Historical Population Register with one and the same user interface. The Intermediate Data Structure (IDS) addresses the needs of statistically oriented users who have a source-critical attitude to all the events about a person (Alter & Mandemakers, 2014), but require users to be more computer literate than the average historian — necessary support services are not sufficiently available. Histreg's web pages serve the needs of those who want to browse pedigrees of their ancestors and make it possible to edit contents based on genealogists' expertise. However, it is up to the users to produce representative statistical results without special support. This is somewhat easier in NHDC's Timeline system because cohorts can be constructed quite flexibly, but that system is designed to track the life courses of individuals rather than following large groups of people (Thorvaldsen, Sommerseth, & Holden, 2020). The North Atlantic Population Project (NAPP) files with linked censuses include weights that allow adjustment of statistical biases in the linked files, but only cover two moments in the life cycle (see [https://international.ipums.org/international/linked\\_data\\_details.shtml](https://international.ipums.org/international/linked_data_details.shtml)). Since most baptismal records contain dates of birth, these can also be linked to the 1910 census and to later censuses. The census from 1950 is linked to the central population register from 1964 as part of the transcription process. We are also inspired by the interfaces developed by our sister institutions in Umeå, Lund, Amsterdam, Salt Lake City and Chicoutimi for future solutions.

Computer algorithms for record linkage and the encoding of family relations have been created both at UiT The Arctic University of Norway and the Norwegian Computing Center. The record linkage algorithms are based on similarities in names, birth year, birthplace, occupation, address and relations with family members. In addition, we also check that the events make up a reasonable life course. The algorithms must be adapted to the characteristics of each source in order to keep the error rates low. For example, it is necessary to require stronger similarities for large municipalities such as Oslo, than for smaller municipalities where fewer persons share equal identities.

The main responsibility for the project is UiT The Arctic University of Norway c/o the Norwegian Historical Data Centre (NHDC). Partners in Norway are the Norwegian Archives, Statistics Norway, the Norwegian Computing Centre, the Norwegian Local History Institute, the National Institute of Public Health, Norwegian School of Economics (NHH) and the University of Bergen.

In summary, the Research Council of Norway has granted resources for the establishment of a national, Historical Population Register (HPR), covering the period from 1800 until the modern,

Central Population Register takes over from 1964. The HPR links censuses and vital records etc. in a common database with the main aim to expand it to a national, historical and longitudinal register for researching social and population history during the last couple of centuries. Also, the merging of different sources at the individual level provides new source-critical insights.

## 9 SUMMARY AND CONCLUSION

The microdata versions of Norwegian historical individual level sources have allowed a wide range of topics within population and social history to be studied more extensively and in more detail. Previously, most publications were either local community studies or using aggregates open to the well-known fallacies. Now, we have a wide range of topics in doctoral dissertations, master theses and articles, and the variety can be illustrated with some examples:

- Infant mortality in Asker and Bærum in the late 18th and early 19th centuries and in multi-ethnic Tana parish from 1840 until World War I.
- Family history in northernmost Norway towards the end of the 19th century and about family structure in 18th century Rendalen.
- How the Spanish flu affected high and low socio-economic groups in Oslo differentially.
- Operationalizing a model for the timing of marriage and illegitimacy in Rendalen, 1750–1900.
- Changes in family structures on the national level using the censuses through the 19th century.
- On migration in Rendalen and south-eastern Norway, tracing internal migration in the provinces surrounding Tromsø and Trondheim during the second half of the 19th century, and the immigration of Jews from 1851 to WWII.
- The struggle of the Sami who trekked with their reindeer between winter pastures in Sweden and summer pastures on the coast of Norway.
- The connection between marriage and ethnicity in Karlsøy 1770–1900 and more generally the classification of ethnicity in Northern Norway.

A few highlights can illustrate the impact of the historical microdata research. We now know that infant mortality levels and childcare practices were inherited over the generations, and that the life conditions of little girls could influence the survival chances of their future children. Premarital sexual relations were usual, and around 1800 up to half of the brides would be pregnant at the altar. We also know that life expectancy decided when means of livelihood became available and thus influenced the timing of marriages. Studies of ethnic marriage patterns have shown resistance against Sami-Norwegian mixed marriages, while the Kvens could function as catalysts to promote inter-ethnic marriage. Primogeniture predominated among the ethnic Norwegians, while the Sami practised ultimogeniture. The family structure varied between different parts of the country, but a general trend was that the relocation of production out of the household to the secondary and tertiary economic sectors from the late 19th century made more women householders, lowered the age at marriage, raised children longer at home and reduced the household size. Emigrant numbers proved to be higher than previously estimated, and when compared to present day international migration, current border restrictions especially limit the emigration of poor people, although this finding may only be valid for urban emigrants.

While priority during the latest decades was given to mortality, migration and family history studies, there is now a development towards studies of social medicine history and to cover the understudied middle decades of the 20th century. Central to this are the use of longitudinal methodology, necessitating record linkage rather than using sources separately, and the building of our Historical Population Register.

Today we have over 10 million manually transcribed person entities from 19th and early 20th century population censuses, and professional transcribers and volunteers have manually transcribed approximately 20 million person entities from the church books. In addition, about 60–80 million person entities transcribed by the previously mentioned genealogy companies, have now been transferred to be used in the Population Register. Our fine-tuned algorithms for record linkage have a success rate of 60–80% for linkage between decennial censuses. For areas where we have a long time-series of

transcribed church books, complete life trajectories have been constructed for the majority of the population.

The most recent development is that the Norwegian Research Council in 2021/2022 granted over 2.5 million Euros for the development of integrated "Historical Registers", which will include the Historical Population Register together with the Property Register and other nominative databases.

In the years to come we will strive to develop optimized automatic systems for preprocessing the 60–80 million data entries, primarily through machine learning (ML) models for standardization and for coding. Given that only a selected group of variables were transcribed by the genealogy companies, we will continue our work on developing ML models for transcription of handwritten text (compare Pedersen et al., 2022). A priority will be given to cause of death data.

Through the successful cooperation with the project "Mapping Human Capital of Nordic Countries", we aim to expand the Historical Population Register with individual level data on education, which today have a national coverage from 1950 and onwards. With the implementation of educational data from a variety of sources, we will be able to go back to the beginning of the 19th century. The recently adopted partner in the Norwegian School of Economics supplies us with individual level income data.

Another aim is to develop online resources, videos and teaching materials as the basis of a resource which we conceive as a 'historical laboratory'. HistLAB, as it has been named, will be offered through our web portal, and researchers, teachers and students from senior secondary schools, colleges and universities will be able to use it along with those data in HPR which are open access.

We believe the establishment of HistLAB will alert established academics, new generations of researchers and students to the existence and scope of our population register and encourage them to use and promote this impressive resource.

## REFERENCES

- Abramitzky, R., Boustan, L. P., & Eriksson, K. (2012). Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration. *American Economic Review*, 102(5), 1832–1856. doi: [10.1257/aer.102.5.1832](https://doi.org/10.1257/aer.102.5.1832)
- Abramitzky, R., Boustan, L. P., & Eriksson, K. (2013). Have the poor always been less likely to migrate? Evidence from inheritance practices during the age of mass migration. *Journal of Development Economics*, 102(C), 2–14. doi: [10.1016/j.jdeveco.2012.08.004](https://doi.org/10.1016/j.jdeveco.2012.08.004)
- Åkerman, S. (1978). Towards an understanding of emigrational processes. In W. H. McNeill & R. S. Adams (Eds.), *Human migration: Patterns and policies* (pp. 287–306). London: Bloomington, IN.
- Alter, G., & Mandemakers, K. (2014). The Intermediate Data Structure (IDS) for longitudinal historical microdata, version 4. *Historical Life Course Studies*, 1, 1–26. doi: [10.51964/hlcs9290](https://doi.org/10.51964/hlcs9290)
- Backer, J. (1961). *Dødeligheten og dens årsaker i Norge 1856–1955* [Mortality and its causes in Norway 1856–1955]. Oslo: Statistics Norway.
- Brändström, A. (1988). The impact of female labour conditions on infant mortality: A case study of the parishes of Nedertorneå and Jokkmokk, 1800–96. *Social History of Medicine*, 1(3), 329–358. doi: [10.1093/shm/1.3.329](https://doi.org/10.1093/shm/1.3.329)
- Brochmann, G., & Kjeldstadli, K. (2008). *A history of immigration: The case of Norway 900–2000*. Oslo: Universitetsforlaget.
- Bull, H. (2000). *Hushold og generasjonsskifter i Rendalen 1762–1900: Ættesamfunnets siste skanse?* [Household and generational change in Rendalen 1762–1900: The last stronghold of the family society?]. Oslo: Universitetet i Oslo. Retrieved from <http://files.christian-asphaug.webnode.no/200000309-55ce756c8c/1762-1900%20H.Bull%20%20C3%86ttesamfunnets%20siste%20skanse%20del%201.pdf>
- Bull, H. H. (2005). Deciding whom to marry in a rural two-class society: Social homogamy and constraints in the marriage market in Rendalen, Norway, 1750–1900. *International Review of Social History*, 50(S13), 43–63. doi: [10.1017/S0020859005002063](https://doi.org/10.1017/S0020859005002063)
- Bull, H. H. (2006). *Marriage decisions in a peasant society: The role of the family of origin with regard to adult children's choice of marriage partner and the timing of their marriage in Rendalen, Norway, 1750–1900* (Doctoral dissertation). Faculty of Humanities, University of Oslo Unipub, Oslo.

- Bygren, L. O., Tinghög, P., Carstensen, J., Edvinsson, S., Kaati, G., Pembrey, M. E., & Sjöström, M. (2014). Change in paternal grandmothers' early food supply influenced cardiovascular mortality of the female grandchildren. *BMC Genetics*, 15. doi: [10.1186/1471-2156-15-12](https://doi.org/10.1186/1471-2156-15-12)
- Coale, A. J., Anderson, B. A., & Härm, E. (1979). *Human fertility in Russia since the nineteenth century*. Princeton: Princeton University Press.
- Drake, M. (1969). *Population and society in Norway, 1735–1865*. Cambridge: Cambridge University Press.
- Dyrvik, S. (1983). *Historisk demografi. Ei innføring i metodane* [Historical demography. An introduction to the methods]. Bergen: Universitetsforlaget
- Dyrvik, S. (1993). Farmers at sea: A study of fishermen in north Norway, 1801–1920. *Journal of Family History*, 18(4), 341–356. doi: [10.1177/036319909301800404](https://doi.org/10.1177/036319909301800404)
- Dyrvik, S. (1997). Spedbarnsdøden i Etne prestegjeld 1716–1900 [Infant mortality in Etne parish 1716–1900]. In G. A. Ersland, E. Hovland & S. Dyrvik (Eds.), *Festskrift til Historisk instituttets 40-års jubileum 1997* (pp. 131–150). Bergen: University of Bergen.
- Dyrvik, S., & Alsvik, M. K. (1987). *Gjennombrøtet for fødselsregulering i Stavanger 1900–1935* [The breakthrough for birth regulations in Stavanger town 1900–1935]. Bergen: University of Bergen.
- Edvinsson, S., & Engberg, E. (2020). A database for the future. Major contributions from 47 years of database development and research at the Demographic Data Base. *Historical Life Course Studies*, 9, 173–196. doi: [10.51964/hlcs9305](https://doi.org/10.51964/hlcs9305)
- Edvinsson, S., Garðarsdóttir, O., & Thorvaldsen, G. (2008). Infant mortality in the Nordic countries, 1780–1930. *Continuity and Change*, 23(3), 457–486. doi: [10.1017/S0268416008006917](https://doi.org/10.1017/S0268416008006917)
- Eide, O. J., & Thorvaldsen, G. (2011). Historisk befolkningsregister: Oversikt over emigrasjonen [The Historical Population Register: An overview of emigration]. *Heimen*, 48(3), 207–228.
- Eliassen, J. (1981). Fine jomfruer, gravide bruder og ugifte mødre. Forholdene i Moss rundt år 1800 [Nice virgins, pregnant brides and unwed mothers. Conditions in Moss town about 1800]. In J. Eliassen & S. Sogner (Eds.), *Bot eller bryllup. Ugifte mødre o gravide bruder i det gamle samfunnet* [A penance or a wedding. Unmarried mothers and pregnant brides in traditional society] (pp. 81–91). Oslo: Universitetsforlaget.
- Engelsen, R. (1983). Mortalitätsdebatten og sosiale skilnader i mortalitet [The mortality debate and social differences in mortality]. *Historisk tidsskrift*, 62(2), 161–202.
- Forsdahl, A. (1973). Momenter til belysning av den høye dødelighet i Finnmark fylke: Kan den høye dødelighet i dag være en senfølge av meget dårlige levevilkår i barne- og ungdomsalderen? [Factors shedding light on the high mortality rate in Finnmark county: Can the high mortality rate today be a late consequence of very poor living conditions in childhood and adolescence?]. *Tidsskrift for den Norske Lægeforening*, 93(10), 661–667.
- Fure, E. (1980). *Den ekteskapelige fruktbarhet i Kolbu, ca. 1870–1910. En studie av fruktbarhetsfallets innledningsfase* [Marital fertility in Kolbu, approximately 1870–1910. A study of fertility decline during the introductory phase] (Doctoral dissertation). University of Oslo, Oslo.
- Fure, E. (1986). Gamle i flergenerasjonsfamilier – en seiglivet myte? [Elderly in multi-generational families – a tenacious myth?]. *Historisk tidsskrift*, 1, 16–35.
- Fure, E. (2000). Spedbarnsdødeligheten og sosiale forskjeller i Asker og Bærum 1814–1878. En metode for studier på individnivå [Infant mortality and social differences in Asker and Bærum parishes 1814–1878. A method for studies at the individual level]. *Heimen*, 37(4), 293–304.
- Fure, E. (2002). Is it the mother's health that really matters? Infant mortality in the parish of Asker and Bærum 1814–1878. In H. Sandvik, K. Telste & G. Thorvaldsen (Eds.), *Pathways of the past: Essays in honour of Sølvi Sogner on her 70th anniversary 15. March 2002* (pp. 73–90). Oslo: Novus.
- Gjelseth, M. (2000). *Relasjonsdatabaser som verktøy i en historisk-demografisk studie* [Relational databases as a tool in a historical-demographic study] (Master thesis). University of Oslo, Oslo.
- Gjernes, M. (2004). Dei fyrste jødiske innvandrane i Kristiania [The first Jewish immigrants in Kristiania/Oslo]. *Historisk tidsskrift*, 83(3), 385–416. doi: [10.18261/ISSN1504-2944-2004-03-02](https://doi.org/10.18261/ISSN1504-2944-2004-03-02)
- Haavet, I. E. (1982). *Avvik eller uhell? Ugifte foreldre omkring 1800 – En sosial analyse* [Deviation or accident? Unmarried parents around 1800 – A social analysis] (Master thesis). University of Bergen, Bergen.
- Hajnal, J. (1965). European marriage patterns in perspective. In D. V. Glass & D. E. C. Eversley (Eds.), *Population in history. Essays in historical demography* (pp. 101–144). London: Edward Arnold.
- Halvorsen, B., & Indseth, K. (1875). *Befolkningsutviklingen i Ullensaker 1733–1845, en demografisk undersøkelse* [Population development in Ullensaker parish 1733–1845, a demography] (Doctoral dissertation). University of Oslo, Oslo.

- Hubbard, W. H. (1995). Historical demography. In W. H. Hubbard, J. E. Myhre, T. Nordby & S. Sogner (Eds.), *Making a historical culture: Historiography in Norway* (pp. 311–324). Oslo: Scandinavian University Press.
- Kjeldstadli, K. (Ed.). (2003). *Norsk innvandringshistorie* [Norwegian immigration history] (Vols. I–III). Oslo: Pax.
- Knudtsen, M. S. (1997). *Fra frelse til helse: Spedbarnsdødelighet og omsorgssyn i Norge ca 1700–1830 med særlig vekt på forholdene i Vår Frue sokn, Trondheim* [From salvation to health: Infant mortality and views on care in Norway approximately 1700–1830 with special emphasis on the conditions in Our Lady parish, Trondheim] (Doctoral dissertation). Norges teknisk-naturvitenskapelige universitet, Trondheim.
- Koren, E. (1979). *Utvandringen fra Ullensaker 1867–99: En sosialhistorisk undersøkelse* [The emigration from Ullensaker municipality 1867–99: A social history study] (Master thesis). University of Oslo, Oslo.
- Kovacevic, M. (2020). *Tuberculosis and society in Tromsø 1878–1920. An epidemiological study of tuberculosis mortality within societal differences* (Master thesis). UiT The Arctic University of Norway, Tromsø.
- Langholm, S. (1974). Historie på individnivå. Omkring Ullensaker-undersøkelsen — et mikrohistorisk eksperiment [History at the individual level. About the Ullensaker survey – a microhistorical experiment]. *Historisk tidsskrift*, 53, 243–272.
- Langholm, S. (1975). Short-distance migration, circles and flows: Movement to and from Ullensaker according to the population lists of 1865. *Scandinavian Economic History Review*, 23(1), 36–62. doi: [10.1080/03585522.1975.10407804](https://doi.org/10.1080/03585522.1975.10407804)
- Larsen, B. (2008). *Giftermål og etnisitet. Samer, nordmenn og ekteskap i Karlsøy 1770–1900* [Marriage and ethnicity. Sami, Norwegians and marriages in Karlsøy parish 1770–1900] (Doctoral dissertation). University of Tromsø, Tromsø. Retrieved from <https://hdl.handle.net/10037/7059>
- Lindbekk, K., & Opach, T. (2020). Befolkningsutvikling i masseutvandringens tid Trøndelag 1866–1914 [Population development in a period of mass emigration Trøndelag 1866–1914]. *Heimen*, 57(1), 23–46. doi: [10.18261/issn.1894-3195-2020-01-03](https://doi.org/10.18261/issn.1894-3195-2020-01-03)
- Lindbekk, K. E. (2017). *Fast og tilstrekkelig levebrød. Livsløpene til trøndere født i 1855 og foreldrene deres 1815–1910* [Permanent and adequate livelihood. The life courses of people born in Trøndelag 1855 and their parents 1815–1910] (Doctoral dissertation). University of Oslo, Oslo.
- Linge, T. (1977). *Befolkningsutviklinga i Nordal 1737–1865* [Population development in Nordal parish 1737–1865] (Master thesis). University of Oslo, Oslo.
- Lithell, U.-B. (1988). Childcare — a mirror of women's living conditions: A community study representing 18th and 19th century Ostrobothnia in Finland. In A. Brändström & L.-G. Tedebrand (Eds.), *Society, health and population during the demographic transition* (pp. 91–108). Umeå: Umeå University.
- Mamelund, S.-E. (2001). Mortality and life expectancy in Rendalen and Norway 1770–1900: Period and cohort perspectives. In L.-G. Tedebrand & P. Sköld (Eds.), *Nordic demography in history and present-day society* (pp. 201–231). Umeå: Umeå University.
- Mamelund, S.-E. (2006). A socially neutral disease? Individual social class, household wealth and mortality from Spanish influenza in two socially contrasting parishes in Kristiania 1918–19. *Social Science & Medicine*, 62(4), 923–940. doi: [10.1016/j.socscimed.2005.06.051](https://doi.org/10.1016/j.socscimed.2005.06.051)
- McCaa, R. (1989). Populate. A microcomputer projection package for aggregative data applied to Norway, 1736–1970. *Annales de Démographie Historique*, 287–298. doi: [10.3406/adh.1989.1747](https://doi.org/10.3406/adh.1989.1747)
- Mølmann, N. P. (2004). *Spedbarnsdødeligheten i en ishavsbys: Hammerfest 1801–1900* [Infant mortality in an Arctic town: Hammerfest 1801–1900] (Master thesis). University of Oslo, Oslo.
- Mørkhagen, S. (2009). *Farvel Norge: Utvandringen til Amerika 1825–1975* [Farewell Norway: The emigration to America 1825–1875]. Oslo: Gyldendal.
- Pedersen, A. G. (2007). Dødelighet og dødsårsaker [Mortality and causes of death]. In R. R. Bore (Ed.), *På liv og død. Helsestatistikk i 150 år* [Life and death. Health statistics during 150 years]. Oslo: Statistics Norway.
- Pedersen, B.-R., Holsbø, E., Andersen, T., Shvetsov, N., Ravn, J., Sommerseth, H. L., & Bongo, L. A. (2022). Lessons learned developing and using a machine learning model to automatically transcribe 2.3 million handwritten occupation codes. *Historical Life Course Studies*, 12, 1–17. doi: [10.51964/hlcs11331](https://doi.org/10.51964/hlcs11331)
- Pitkänen, K. (1983). Infant mortality decline in a changing society. *Finnish yearbook of population research*, 21, 46–74. doi: [10.23979/fypr.44768](https://doi.org/10.23979/fypr.44768)

- Quaranta, L., Broström, G., van Dijk, I., Donrovich, R., Edvinsson, S., Engberg, E., . . . Sommerseth, H. L. (2017). *Intergenerational transfers of infant mortality in historical contexts: A comparative study of five European populations*. Paper presented at the Population Association of America (PAA), Chicago, IL, USA.
- Quaranta, L., & Sommerseth, H. L. (2018). Introduction: Intergenerational transmissions of infant mortality using the Intermediate Data Structure (IDS). *Historical Life Course Studies*, 7, 1–10. doi: [10.51964/hlcs9288](https://doi.org/10.51964/hlcs9288)
- Randsborg, H. B. (1979). *Befolkningsutviklingen i Bø i Telemark: 1727–1815* [Population development in Bø parish in Telemark county: 1727–1815] (Master thesis). University of Oslo, Oslo.
- Ruggles, S. (1994). The transformation of American family structure. *American Historical Review*, 99(1), 103–128. doi: [10.1086/ahr/99.1.103](https://doi.org/10.1086/ahr/99.1.103)
- Sogner, S. (1978). Familie, husstand og befolkningsutvikling [Family, household and population development]. *Heimen*, XVII.
- Sogner, S. (1979). *Folkevekst og flytting. En historisk demografisk studie i 1700-årenes Øst-Norge* [Population growth and relocation. A historical demographic study of 18th-century eastern Norway] (Doctoral dissertation). Universitetsforlaget, Oslo
- Sogner, S. (1990). *Far sjøl i stua og familien hans: Trekk fra norsk familiehistorie før og nå* [The father himself in the living room and his family: Aspects of Norwegian family history before and now]. Oslo: Universitetsforlaget.
- Sogner, S. (2000). A case study of women's role and infant mortality. In S. Sogner & G. Hagemann (Eds.), *Women's politics and women in politics. In honour of Ida Blom (45–67)*. Oslo: Cappelen Akademisk Forlag.
- Sogner, S. (2009). The Norwegian stem family: Myth or reality? In A. Fauve-Chamoux & E. Ochiai (Eds.), *The stem family in Eurasian perspective. Revisiting house societies, 17th–20th centuries* (pp. 151–172). Bern: Peter Lang.
- Sogner, S. (2012). "Og skuta lå i Amsterdam ...": *Et glemt norsk innvandrersamfunn i Amsterdam 1621–1720* ["And the sail ship was in Amsterdam ...": A forgotten Norwegian immigrant community in Amsterdam 1621–1720]. Oslo: Aschehoug.
- Sogner, S. (2016). Historical demography in Norway 1960–2010. In A. Fauve-Chamoux, I. Bolovan & S. Sogner (Eds.), *A global history of historical demography: Half a century of interdisciplinarity* (pp. 499–512). Bern: Peter Lang.
- Sogner, S., Fure, E., & Randsborg, H. B. (1984). *Fra stua full til tobarnskull: Om nedgangen i barnetall i norske familier i de siste 200 år, med særlig vekt på perioden 1890–1930* [From a housefull to parity two: On the decline in child numbers during the last 200 years, emphasizing 1890–1930]. Oslo: Universitetsforlaget.
- Sogner, S., & Thorvaldsen, G. (2001). *Surnames as proxies for place of origin in the 1801 census for Norway*. Paper presented at the Nordic Demographic Symposium, Tjøme, Norway.
- Solli, A. (1995). *Individ – hushald – samfunn: Fisk og jord som grunnlag for hushaldsetableringa i Norge kring år 1800* [Individual - household - society: Fish and soil as a basis for the establishment of households in Norway around the year 1800] (Master thesis). University of Bergen, Bergen. Retrieved from <https://hdl.handle.net/1956/6286>
- Solli, A. (2004). *Livsløp – familie – samfunn: Endring av familiestrukturar i Norge på 1800-talet* [Life course - family - society: Changing family structures in Norway in the 19th century] (Doctoral dissertation). University of Bergen, Bergen. Retrieved from <https://hdl.handle.net/1956/3061>
- Sommerseth, H. L. (2003). *Spedbarnsdødeligheten i Tana i perioden 1840–1914* [Infant mortality in Tana in the period 1840–1914] (Master thesis). University of Tromsø, Tromsø.
- Sommerseth, H. L. (2006). Fikk spedbarnet morsmelk? Spedbarnsernæringen i Finnmark og en problematisering av hvordan denne påvirket det relativt høye dødelighetsnivå i Tana i perioden 1840–1914 [Did the infant receive breast milk? The infant nutrition in Finnmark county and problematizing how this affected the relatively high mortality rate in Tana parish during the period 1840–1914]. *Heimen*, 43(1), 3–14.
- Sommerseth, H. L. (2011a). The effect of ethnicity and economy upon intergenerational coresidence: Northern Norway during the last part of the nineteenth century. *Journal of Family History*, 36(3), 263–285. doi: [10.1177/0363199011406634](https://doi.org/10.1177/0363199011406634)
- Sommerseth, H. L. (2011b). Endring i samisk og norsk husholdsstruktur: Nord-Troms og Finnmark i perioden 1865 til 1900 [Changing Sami and Norwegian household structure: North Troms and Finnmark counties during the period 1865 to 1900]. *Historisk tidsskrift*, 90(1), 33–61 doi: [10.18261/ISSN1504-2944-2011-01-03](https://doi.org/10.18261/ISSN1504-2944-2011-01-03)

- Sommerseth, H. L. (2018). The intergenerational transfer of infant mortality in Northern Norway during the 19th and early 20th centuries. *Historical Life Course Studies*, 7, 69–87. doi: [10.51964/hlcs9284](https://doi.org/10.51964/hlcs9284)
- Sommerseth, H. L., & Thorvaldsen, G. (2016). Masculine responsibility across generations: Living arrangements in a Norwegian parish around 1900. *The History of the Family*, 21(1), 21–37. doi: [10.1080/1081602X.2015.1069204](https://doi.org/10.1080/1081602X.2015.1069204)
- Sommerseth, H. L., & Walkout, E. C. (2019). Deaths in a city: A view from the 19th century church registers in Norway. In E. Glavatskaya, G. Thorvaldsen, G. Fertig & M. Szoltysek (Eds.), *Nominative data in demographic research in the East and the West* (pp. 185–201). Ekaterinburg: Ural University Press. Retrieved from [https://elar.uifu.ru/bitstream/10995/73898/1/978-5-7996-2656-3\\_2019.pdf](https://elar.uifu.ru/bitstream/10995/73898/1/978-5-7996-2656-3_2019.pdf)
- Statistics Norway. (1980). *Folketeljinga 1801. Ny bearbeiding* [Population census 1801. Reprocessed]. Oslo: Central Bureau of Statistics. Retrieved from [https://www.ssb.no/a/histstat/nos/nos\\_b134.pdf](https://www.ssb.no/a/histstat/nos/nos_b134.pdf)
- Statistics Norway. (1995). *Historisk statistikk 1994* [Historical statistics 1994]. Oslo: Statistics Norway. Retrieved from [https://www.ssb.no/a/histstat/nos/nos\\_c188.pdf](https://www.ssb.no/a/histstat/nos/nos_c188.pdf)
- Sunde, R. (2001). *Vikjer ved fjorden, vikjer på prærien: Ein demografisk-komparativ studie med utgangspunkt i Vik i Sogn* [Vik people by the fjord, Vik people on the prairie: A demographic-comparative study with a starting point in Vik parish in Sogn county] (Doctoral dissertation). Høgskulen i Sogn og Fjordane, Sogndal.
- Szolysek, M., Ogórek, B., & Gruber, S. (2021). Global and local correlations of Hajnal's household formation markers in historical Europe: A cautionary tale. *Population Studies*, 75(1), 67–89. doi: [10.1080/00324728.2020.1832252](https://doi.org/10.1080/00324728.2020.1832252)
- Thorvaldsen, G. (1977). Bruk av elektronisk databehandling i bearbeidelsen av lokalhistoriske kilder [Using electronic data processing in the processing of local history sources]. *Heimen*, XVII, 493–498.
- Thorvaldsen, G. (1978). *Registreringssentral for historiske data* [The Norwegian Historical Data Centre]. Retrieved from <https://rhd.uit.no/indexeng.html>
- Thorvaldsen, G. (1979). *Registreringssentral for historiske data. Rapport fra prøveprosjektet 1/8 1978 til 28/2 1979* [The Norwegian Historical Data Centre. Report from the pilot project 1.8.1978 to 28.2.1979]. Institute of Social Sciences, University of Tromsø: Tromsø.
- Thorvaldsen, G. (1995). *Migrasjon i Troms i annen halvdel av 1800-tallet. En kvantitativ analyse av folketellingene 1865, 1875 og 1900* [Migration in Troms province in the second half of the 19th century. A quantitative analysis of censuses 1865, 1875 and 1900] (Doctoral dissertation). Universitetet i Tromsø, Tromsø. Retrieved from <https://hdl.handle.net/10037/1390>
- Thorvaldsen, G. (1996). *Håndbok i registrering og bruk av historiske persondata* [Manual for registration and use of historical personal data]. Oslo: Tano Aschehoug.
- Thorvaldsen, G. (1998). Historical databases in Scandinavia. *The History of the Family. An International Quarterly*, 3(3), 371–383. doi: [10.1016/S1081-602X\(99\)80252-8](https://doi.org/10.1016/S1081-602X(99)80252-8)
- Thorvaldsen, G. (2002). Rural infant mortality in nineteenth century Norway. In S. Sogner (Ed.), *Studies in mortality decline. Report from the Mortality Studies Group* (75–113). Oslo: Centre for Advanced Study. doi: [10.3384/hygiea.1403-8668.023175](https://doi.org/10.3384/hygiea.1403-8668.023175)
- Thorvaldsen, G. (2011). Using NAPP census data to construct the Historical Population Register for Norway. *Historical Methods*, 44(1), 37–47. doi: [10.1080/01615440.2010.517470](https://doi.org/10.1080/01615440.2010.517470)
- Thorvaldsen, G. (2014). Religion in the census. *Social Science History*, 38(1–2), 203–220. doi: [10.1017/ssh.2015.16](https://doi.org/10.1017/ssh.2015.16)
- Thorvaldsen, G. (2018a). *Censuses and census takers: A global history*. London: Routledge.
- Thorvaldsen, G. (2018b). Emigrants in the Historical Population Register of Norway. *Journal of Migration Research*, 4(2), 264–288. doi: [10.1163/23519924-00402003](https://doi.org/10.1163/23519924-00402003)
- Thorvaldsen, G. (2019). Internal migration in 19th and 20th century Norway. An overview 1865 to 1960. In E. Glavatskaya, G. Thorvaldsen, G. Fertig & M. Szoltysek (Eds.), *Nominative data in demographic research in the East and the West* (pp. 166–184). Ekaterinburg: Ural University Press. Retrieved from [https://elar.uifu.ru/bitstream/10995/73898/1/978-5-7996-2656-3\\_2019.pdf](https://elar.uifu.ru/bitstream/10995/73898/1/978-5-7996-2656-3_2019.pdf)
- Thorvaldsen, G., Sommerseth, H. L., & Holden, L. (2020). Anvendelser av Norges historiske befolkningsregister [Interfaces to Norway's Historical Population Register]. *Heimen*, 57(3), 230–243. doi: [10.18261/issn.1894-3195-2020-03-05](https://doi.org/10.18261/issn.1894-3195-2020-03-05)

HISTORICAL LIFE COURSE STUDIES  
VOLUME 12 (2022), published 07-10-2022

# Historical Life Courses and Family Reconstitutions

## The Scientific Impact of the Antwerp COR\*-Database

Paul Puschmann

Radboud University Nijmegen, the Netherlands

Hideko Matsuo

KU Leuven, Belgium

Koen Matthijs

KU Leuven, Belgium

### ABSTRACT

The Antwerp COR\*-database is a longitudinal micro-level database, which covers all entries from individuals whose last names started with the letters COR (and individuals who shared at some moment in time a household with a COR\*-person) from the population registers and the vital registration of births, marriages and deaths for the 19th- and early-20th-century Antwerp district in Flanders, the northern Dutch-speaking part of Belgium. As such the database allows the reconstruction of historical life courses and families, and the analysis of key demographic characteristics and developments regarding marriage, fertility, migration, social mobility, health, mortality and longevity, as well as their interplay within and across households, families and generations. After a short description of the source material and the construction of the database, a review of the literature based on the database is presented in order to provide the reader with an encompassing overview of the research that has been carried out with this database and the knowledge and insights it has generated since its first release in 2010. The article ends with a discussion of potential pathways for future research, including new topics, and future extension of the database through citizen science projects.

**Keywords:** Antwerp COR\*-Database, Longitudinal data, Life courses, Family reconstitution, Population registers, Civil registry

e-ISSN: 2352-6343  
PID article: <https://doi.org/10.51964/hlcs12914>

© 2022, Puschmann, Matsuo, Matthijs

This open-access work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/by/4.0/), which permits use, reproduction & distribution in any medium for non-commercial purposes, provided the original author(s) and source are given credit. See <http://creativecommons.org/licenses/>.

## 1 INTRODUCTION

Thanks to the efforts of Adolphe Quetelet, the newly-independent Kingdom of Belgium developed in the 1840s a population administration, which would set an example for many other European countries. The Central Committee for Statistics — under the leadership of Quetelet himself — started in 1846 a population register at the municipality level with a nation-wide coverage. Census sheets were copied down in new books, covering cross-sectional data by household. Each double page referred to one address, and each line of the population register consisted of one individual, with the head of the household on top — usually a man —, followed by the spouse, children and other household members, including kin — e.g., resident siblings of the head of the household and/or his wife — and non-kin, for instance, lodgers and domestic servants. For each individual, core characteristics were entered, including the relation to the head of the household (spouse, child, brother, sister, etc.), sex, birth date, birth place, marital status and occupation. So far the register was not more than a copy of the census. However, in the decade to come — until a new census was taken — all mutations in the household were meticulously registered: the birth of children, the death and in- and out-migration of family members, as well as changes in marital status. Since all life course events were dated in this way, the population register was a living document that kept track of a number of demographic changes at the individual and the household level. Together with the civil registry that had been installed in the 1790s under Napoleon, the population registers offered the Belgian authorities a superb administrative tool for monitoring its population (Gutmann & Van de Walle, 1978; Oris & Ritschard, 2014).

Figure 1 Example of a population register from Antwerp city — 1890–1900

NAA M.	VOORNAMEN.	BEROEP, BURGERLIJKE STAND en graad van VERWANTSCHAP met het hoofd des HUISGEZINS.	GEBORTEPLAATS, JAAR, MAAND en DATUM	INTREDE, INWENIGSE VERANDERINGEN EN VERTREK.		AANMERKINGEN.
				DATUM VAN INKOMING in de STAD.	WOONPLAATS en VERBLIJF.	
1	Cortoos	Stans	1871	1-2-98	Deurne 27 5 6 8 98 Charleston Co- 4 98	
2	Hens	Karin	1872	1-2-98	From. St. Borgankouke 22 11 18 98	
3	Cortoos	Petrus	1898	16-2-	Overleiden 27 4 98	
4	Cortoos	Petrus	1899	1-2-	Green St. Overleiden 27 5 18 99	

Source: Scan from FelixArchief — <https://felixarchief.antwerpen.be/archievenoverzicht/87729>

Notwithstanding the very rich data sources, historical demographers in Belgium — and elsewhere in Europe — relied until far into the 20th century almost exclusively on aggregate statistics — mostly from censuses and demographic yearbooks — or they applied family reconstruction on birth and marriage certificates, as outlined by Michel Fleury and Louis Henry (1956), neglecting the rich individual level data provided by the population registers. It was only after Myron Gutmann and Etienne van de Walle (1978) published their article "New Sources for Social and Demographic History: The Belgian Population Registers" that an interest arose in the use of these superb data sources. Especially George Alter's (1988) "Family and the Female Life Course, the Women of Verviers", which showed the power of the population registers by applying event history analysis to marriage and fertility data, caused a paradigm shift in the field. The applied methodology, which Alter described carefully, was superior to the family reconstructions, as it was able to deal with incomplete life courses — i.e., censoring — by taking the time at risk into account, and avoided biases resulting from the neglect of families with incomplete information, often due to migration (Alter, 2020; Ruggles, 1999).

Alter's monograph marked the start of an increasing number of historical demographic studies based on individual level data from Belgian population registers and vital registers. Studies appeared on marriage, fertility, migration, mortality, and social mobility, as well as their relationships (e.g., Alter & Oris, 2005; Neven, 2003; Oris, 1996). These studies mainly focused on Wallonia — the southern French-speaking part of Belgium, a region that experienced early industrialization and urbanization —, thereby largely neglecting Flanders, the northern, Dutch-speaking part of the country.

## 2 THE CREATION AND DEVELOPMENT OF THE ANTWERP COR\*-DATABASE

Inspired by the above mentioned studies on Wallonia, the development of the Historical Sample of the Netherlands (Mandemakers & Kok, 2020), as well as the construction of Swedish historical demographic databases at Umeå University (Engberg & Edvinsson, 2020; Westberg, Engberg, & Edvinsson, 2016) and Lund University (Bengtsson & Dribe, 2021; Dribe & Quaranta, 2020), the idea arose to build a high quality historical demographic database on Flanders. Under the lead of Koen Matthijs, who had acquired extensive experience on the collection and analysis of vital registration data (Matthijs, 2002, 2003a), members of the research group Family and Population Studies (FaPOS) of the KU Leuven, started a pilot project on the city of Antwerp in order to test whether the envisioned data collection procedure was feasible (Matthijs & Moreels, 2010; Van Baelen, 2007).

As it was unrealistic to collect the socio-demographic information on all individuals and families who lived in Antwerp city and beyond during the second half of the 19th and early 20th century, it was necessary to select an efficient sampling strategy. Following the data collection principle of the French TRA-Project (Bourdieu, Kesztenbaum, & Postel-Vinay, 2014), a letter sample was chosen. In practice, all entries from birth, marriage and death certificates, as well as from the population registers of individuals whose family names started with the letters "COR" (e.g., Cordon, Cornelissen, Coremans, Corlui, Cornet) were collected, stored, cleaned and linked. Last names with this letter combination were chosen as investigations by the KU Leuven Department of Linguistics showed that individuals with these last names were equally distributed over Flanders and representative of the Flemish population in terms of a number of socio-demographic indicators, including SES and migration status (Moreels & Matthijs, 2011; Van Baelen, 2007). The latter was especially important since Antwerp turned into the fastest growing and most populous city in Belgium, as a result of large-scale in-migration in the course of the 19th century (Kruithof, 1964; Winter, 2009). Equally important was the fact that COR\* family names were not linguistically exclusive, making sure that immigrant families from various countries were included. More than 7% of the individuals in the database is indeed of foreign descent, with a majority of immigrants from the neighboring Netherlands, Germany and France, but the database contains also individuals who were born in faraway countries, such as Australia, Egypt, Japan, the United States and Russia.

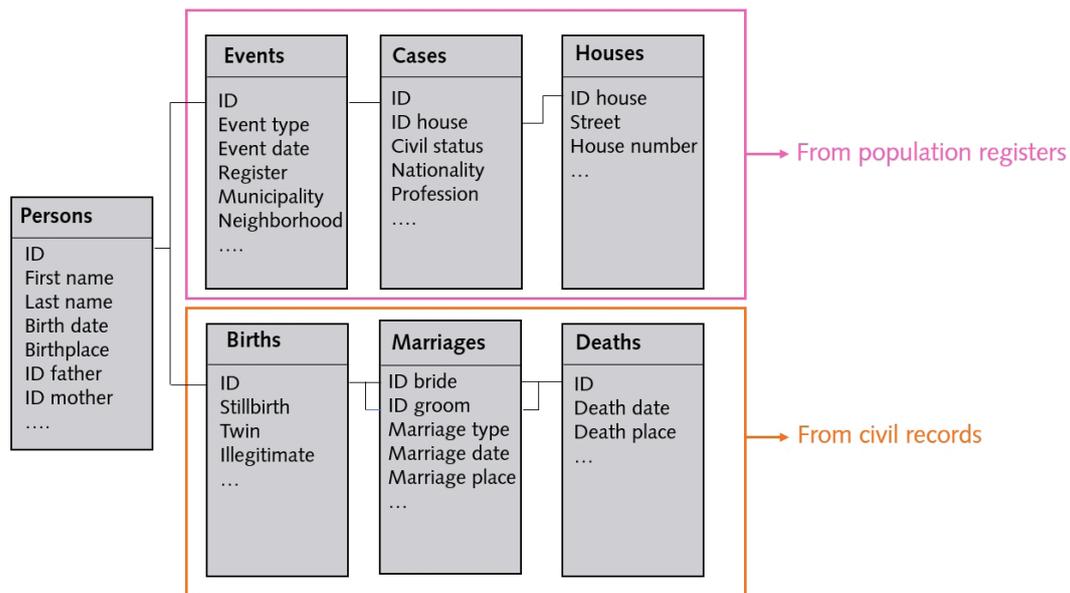
Working with a letter sample had several advantages. First, it simplified the data collection and therefore increased the reliability and quality of the data, as data collectors only had to search for individuals based on their last name instead of a whole list of criteria. Second, the reconstruction of life courses and families was facilitated by selecting people based on their family name, as it made sure that throughout the whole research area and period pieces of information from the same families and individuals were selected. Third, this sampling strategy made it easier to deal with lost and incomplete information, which was imperative in the highly mobile urban society of Antwerp, in which people moved frequently within and across municipalities, and not all migrations might have been properly registered. Instead of following the lives of individuals through time and space, which would have meant that data collectors would have to follow individuals from one municipality register to the next over time, sometimes being confronted with "dead ends", all data entries on research persons were collected following the "vacuum cleaning method": all pieces were collected independently by multiple data collectors. The life courses were later reconstructed by putting all pieces together by way of record linkage (Matthijs & Moreels, 2010; Van Baelen, 2007).

The Antwerp COR\*-database covers data from 1846 — the year when the first nation-wide population registers were issued — until 1920, the end of the 1910 register. The end point was pragmatically chosen as later population registers at the time of the data collection were not yet accessible due to privacy regulations.<sup>1</sup> The pilot project on the city of Antwerp turned out to be successful and therefore it was decided to enlarge the area of data collection to Antwerp's suburbs and the rural municipalities of the larger Antwerp district. This resulted in the 2010 release of the Antwerp COR\*-database, which forms the basis of most historical-demographic studies on the 19th- and early-20th-century Antwerp district.<sup>2</sup>

1 Note that there are recent developments in Belgium privacy law, shortening the individual privacy protection from 100 years (birth certificates) to 75 years (marriage certificates) and 50 years (death certificates).

2 For an extensive description of the construction and composition of the Antwerp COR\*-database, see Matthijs and Moreels (2010) and Van Baelen (2007).

Figure 2 The structure of the 2010 release of the Antwerp COR\*-database



An important advantage of the database is that it allows not only to track family members residing within the household, but also those residing outside the household, a feature that is usually lacking in databases that consist of intergenerational historical life course data based on a random sample from the population under study (Van Baelen & Matthijs, 2007).

In the context of EU Horizon 2020 Marie Skłodowska-Curie project "Methodologies and data mining techniques for the analysis of Big Data based on Longitudinal Population and Epidemiological Registers" (LONGPOP, MSCA-ITN 676060), the COR\*-database was transformed into the Intermediate Database Structure (IDS; Alter & Mandemakers, 2014), after a systematic evaluation of the original methodologies for the address-based reconstruction of households based on a set of criteria, and the detailed geocoding of the historical database. This resulted in the 2020 IDS-release of the COR\*-database, a release that offers plenty of opportunities for comparative analysis (Jenkinson, Anguita, Paiva, Matsuo, & Matthijs, 2020).

### 3 LITERATURE REVIEW

Since the release of the Antwerp COR\*-database in 2010 (Centre of Sociological Research, research group Family and Population Studies, KU Leuven, 2010) various empirical studies have been conducted on the data. In this section we summarize the most important findings, by situating and discussing them in the context of the broader historiographical debates. The focus will be on articles, chapters and working papers, but we also report findings from PhD theses and a couple of bachelor papers and master theses. This review is organized thematically. We start with partner choice and (re)marriage, subsequently we deal with fertility and bridal pregnancies, out of wedlock fertility and extra-pair-paternity, morbidity and mortality, migration and social inclusion, and social stratification and social mobility.

#### 3.1 MARRIAGE AND PARTNER CHOICE

In the 19th and early-20th century marriage was a key transition in the life course, marking the entry into adulthood. Usually it was coupled with a move from the family of orientation to the family of procreation (Puschmann, 2020a). Western Europe was characterized by the so-called European Marriage pattern of late and non-universal marriage. Couples were only expected to get married once they were financially independent and had the means to set up their own household (Hajnal, 1965, 1982).

Katherine Lynch (1991) found an urban variant of this pattern, marked by even higher ages at marriage and larger shares of singles than in the Western European countryside. Sarah Moreels and Koen Matthijs (2011) observed the urban variant of the European marriage pattern also for the city of Antwerp. Moreover, they found differences with other Flemish cities. Whereas in the industrializing textile cities of Aalst and Ghent, ages at marriages started to decrease from the middle of the 19th century on — signaling the granulation of the Western European marriage pattern — ages at first marriage in the port city of Antwerp increased until about 1890. This was to a large degree explained by heavy urban in-migration. Further research confirmed that both internal and international migrants in Antwerp — as well as in other major European port cities — were less likely to marry and did so at significantly higher ages than the native urban population. The highest ages at first marriage and the highest likelihood to remain single were found among the international migrants (Puschmann, Grönberg, Schumacher, & Matthijs, 2014a; Puschmann, Van den Driessche, Grönberg, Van de Putte, & Matthijs, 2015; Puschmann, Van den Driessche, Matthijs, & Van de Putte, 2012; Puschmann, Van den Driessche, Matthijs, Van de Putte, 2016a; Schumacher, Matthijs, & Moreels 2013). These results suggest that migrants had a hard time in becoming part of urban mainstream society, a finding to which we return in the section on migration and social inclusion.

Next to rural-urban residence and migration status, social status also played a role in the access to marriage, in the sense that the higher classes married later than the lower ones, suggesting that there was more at stake for the elite when it came to marriage partner choice (Moreels & Matthijs, 2011), but it might have been also a result of a longer period of education and training (Caron, Neyrinck, Dillon, Matthijs, 2017). Last but not least, for the elite women it might have been simply more difficult to find a spouse with an equal (or higher) social status, as the group of elite men was relatively small, while these men were due to their power and wealth attractive to women from all social layers, and marrying downwards was often less of an issue for men than for women. Another group with relatively high marriage ages were the sons of farmers, which has been interpreted by Marianne Caron et al. (2017) as a sign of land saturation. Due to strong population growth and innovations in agriculture, everywhere in Europe it became indeed increasingly difficult for young men and women to obtain their own farm or to find work as agricultural laborers, resulting in significant out-migration (Mönkediek, Kok, & Mandemakers, 2016).

Marianne Caron et al. (2017) also found evidence that sibship composition played a role in getting access to the marriage market. In this regard, siblings could act as competitors, but they could also improve each other's chances of finding a partner and getting married. Moreover they also influenced the timing of marriage. It was found that having older siblings extended individuals' waiting time until marriage, while having younger siblings shortened that time. The negative effect of having older siblings became stronger at higher ages. Next, gender effects were found in the sense that for both women and men a recent marriage of a sibling of the same sex delayed marriage, while for marriages of siblings of the opposite sex an acceleration effect was found. While the first result was interpreted in terms of resource dilution, i.e., the parental resources were temporarily exhausted, the latter might point at the positive influence of the extension of the social network through a brother-in-law or a sister-in-law. The total number of siblings did not influence the age at marriage.

Marriage dates can also be used as an indicator of broader social changes. Marriage seasonality can, for example, be used as a proxy for secularization as marriages during Lent and Advent were "forbidden" or discouraged by the Catholic Church. Hideko Matsuo and Koen Matthijs (2018) assessed the extent and evolution of church control through the development of Lent and Advent marriages. They also examined this for fertility by studying whether conceptions occurred during the religious "closed periods", as sex was also forbidden during these religious spells, especially during Lent. Their study showed declining compliances of religious rules on marriage and to some extent on conception, indicating a secularization trend. Different underlying mechanisms existed: marriages were more influenced by social control than conceptions, and the higher the birth order, the lower the level of compliance. Socio-economic, cultural and demographic variables in the COR\*-database allowed to examine who was more, or less likely to adhere to church rules. For marriage, non-compliances were largely observed among urban citizens and older brides, which increased in later periods (early 20th century). Compliance, by contrast, was found mainly among elite bridegrooms. For conceptions, non-compliance was found among literate women, while again compliance with the church rules was found among elite bridegrooms.

## 3.2 FERTILITY DECLINE

In the course of the 19th century the fertility transition took off in Western Europe. Scholars agree that fertility decline during the first demographic transition was realized through "traditional" techniques: the calendar method and coitus interruptus. These were not particularly safe for individuals and couples, but highly effective on the macro-level. It is also clear that the decline of mortality, and in particular infant and child mortality, was an important contributor to fertility decline (Haines, 1998), because it led *ceteris paribus* to an increase in average family size, creating a first incentive to limit the number of live births. Next, there were economic motives to limit the number of offspring. As a result of bans on child labor and the introduction of compulsory schooling, children started to contribute only much later in life to the family income, while at the same time raising them became more expensive. Consequently, the opportunity costs of children increased, turning around intergenerational wealth flows, through which it became rational to limit fertility (Caldwell, 1976, 2005). Moreover, as children demanded ever increasing investments in human capital, it became logical to invest in child quality instead of quantity (Becker, 1960; Spolaore & Wacziarg, 2021). However, there were also cultural and religious factors that prevented couples from actually practicing birth control, when this had become economically rational and feasible (Engelen, 1997). In that sense, to practice birth control couples had to be "ready, willing and able", as Ansley Coale (1973) concluded.

While the complex interplay of economic motivations and cultural and religious impediments — that has been studied extensively within the European Fertility Project (see Coale & Watkins, 1986) — explains an important part of the complex history of fertility decline, it has become also clear that this is not the complete story. After all, one would expect that countries that industrialized first were also the first to see their fertility decline. The simple fact that the fertility transition started in France and not in England, shows that the complex puzzle is not yet solved. Over the years, historical demographers have therefore considered other factors and mechanisms, including the role of family systems (e.g., Rotering, 2020) and diffusion mechanisms (i.e., the idea of "communicating communities", see Szepter, 1996). While studies have increased our insights, the historical debate on the fertility decline continues, mainly because the striking geographic and social differences in the onset and progress of the fertility seem not to be explainable by one encompassing theory.

All these complexities in the debate on fertility decline are also observable in Belgium. Belgium was the first country on the European continent to industrialize and it was also one of the forerunners in the fertility decline (Lesthaeghe, 2015). However, if one zooms in, it turns out that there were major geographic and social differences in the timing and pace of fertility decline. In the industrial zones of Wallonia, the decline took off earlier and fertility levels remained below those of Flanders in the course of the 19th and early 20th century. Fertility was also higher among Flemish couples that had similar income and educational levels as French-speaking couples. For the city of Leuven e.g., evidence was found that parity-dependent stopping behavior of Flemish laboring class couples was influenced on the neighborhood level by the presence of francophone couples and couples from the upper classes. These groups adopted fertility control earlier on than Flemish laboring couples. This suggests that diffusion effects played a role in the Belgian fertility decline (Van Bavel, 2004).

Hideko Matsuo and Koen Matthijs (2016) examined the interplay of socio-economic and cultural factors in fertility limitation behavior during the demographic transition, by studying this through the age at which the last child was born. Based on Kaplan-Meier analyses and hazard models, their analysis confirmed that the age at last childbearing declined mainly among the birth cohort of 1860–1888 in comparison to earlier cohorts (1800–1839 and 1840–1859). They identified substantial inter- and intra-cohort differences, driven by cultural and life course factors, such as literacy status, spousal age gaps, witness characteristics, birth seasonality, marital and childbearing ages and birth histories.

Diffusion effects were found as well. Sarah Moreels and Matthijs Vandezande (2012) investigated parity-specific stopping behavior of 747 couples from the city of Antwerp and categorized them into native, migrant and mixed couples. Generally speaking, native couples were the first to practice birth control. Interestingly, migrants who moved as children to Antwerp adopted the stopping behavior of the urban population of Antwerp, whereas immigrants who moved as adults to the city resembled the behavior of the population of origin. The analysis of the mixed couples showed that the origin of the wife was more decisive than that of the husband, suggesting that — at least in this context — women took the lead when it came to the adoption of fertility control.

Reto Schumacher et al. (2013) studied family formation trajectories of migrants and natives in Antwerp and Geneva from a life course perspective, using sequence analysis. They found for Antwerp important differences in the reproductive careers of short- and long-distance migrants, the former being highly traditional (i.e., long childbearing periods, little or no signs of stopping behavior, and a high average number of children ever born), while the latter were forerunners in the fertility transition (i.e., short childbearing periods, signs of stopping, and a lower number of children). Interestingly, for 19th century Geneva, the differences in the reproductive careers of native and migrant women were smaller. This might be related to the type of migrants that those cities attracted. Geneva received mainly skilled migrants, while Antwerp received mainly unskilled laborers (due to its port function), and this applied especially to the short-distance migrants.

### 3.3 BRIDAL PREGNANCIES, OUT-OF-WEDLOCK FERTILITY AND EXTRA-PAIR PATERNITY

While couples were expected to wait with having sex until they were married and to remain faithful to their marriage partner, deviations from those norms did occur. In European pre-industrial societies it was not uncommon for couples to have sex in anticipation of marriage. In fact, there were all kinds of local practices in which unmarried couples had sex in peer-controlled settings. In case the young woman involved got pregnant, the mating partner would voluntarily or under pressure of the parents and/or others, marry her (Kok, Bras & Rotering, 2016). However, from the latter half of the 18th century on, there was a significant increase in the number of out-of-wedlock births in Western European societies, which peaked in the course of the 19th century, before declining by the end of the 19th century, although the timing varied across regions (Kok, 2005; Matthijs, 2001). The phenomenon was mainly, but not exclusively, observed in urban areas under the laboring classes.

Various explanations have been put forward to explain the rise and decline in extra-marital fertility. Edward Shorter (1971, 1973) suggested that it signified an early sexual revolution among laboring class women, resulting from the liberating effect of earning wages in industry. Peter Laslett (1980) claimed that out-of-wedlock fertility increased within so-called "bastardy-prone sub-societies", in which out-of-wedlock fertility was no deviation from the norms and in which women often gave birth to multiple children out of wedlock. However, most studies point to the role of uncertainty — among both women and men — and vulnerability that increased among young women in the wake of industrialization, rural-to-urban migration and large scale urbanization. The lack of a social network, job insecurity, and relatively low wages all weakened the bargaining power of young women vis-à-vis men. At the same time, illegitimacy might have been both a cause and a consequence of vulnerability (Schumacher, Ryczkowska, & Perroux, 2007).

Against this background, Sophie Vries (2019) studied in her bachelor thesis bridal pregnancies and illegitimate births in the Antwerp district. She observed that — in line with the expectations based on other studies — both phenomena increased in the course of the 19th century and declined again in the beginning of the 20th century. She systematically compared the life courses of (1) women who conceived all children within wedlock, (2) women who experienced a bridal pregnancy, and (3) women who gave birth to one or more illegitimate children. The analysis showed that the life course of women with illegitimate children differed much more from women who conceived all children within marriage, than those of women who experienced a bridal pregnancy. The results made assumable that — in line with the research of Reto Schumacher et al. (2007) — out-of-wedlock fertility was mainly a result of vulnerability. The transition from a regional textile center into a port city led to lower incomes, less job stability, and a substantial decline in job opportunities for females which made them more vulnerable, resulting in a rise in illegitimate births. While out-of-wedlock fertility was mainly caused by vulnerability, it did not necessarily make the women even more vulnerable, since many of the women who gave birth to illegitimate children would ultimately marry. An exception to this were the group of international migrant women, as they often gave birth to multiple children out of wedlock, while their likelihoods of remaining single were much higher than among the natives and internal migrant women.

The debate about out-of-wedlock fertility also raised questions about how faithful couples were to each other (Puschmann, 2020b). Thanks to cooperation between historical demographers and geneticists some light has been shed on biological kinship ties in the past, using historical demographic and genealogical data in combination with DNA information. By comparing Y-chromosomal DNA of contemporary individuals who have the same legal ancestor, it can be determined whether in previous generations women had children with men who were not the legal fathers, as the non-recombining part of Y-chromosome is transmitted unchanged from fathers to sons (Larmuseau, 2021; Larmuseau et al., 2017).

Making use of the Antwerp COR\*-database and other historical demographic and genealogical data, Maarten Larmuseau et al. (2013) found that in Flanders extra-pair paternity was rare in the past. Linking historical demographic data with contemporary genealogical data — allowing generational linkage and by doing so, covering multidisciplinary fields — was one of the important contributions COR\*data has made in this respect. Extra pair paternity rates were around 1–2% per generation, much lower than previous studies had suggested. Further research (Larmuseau, 2021; Larmuseau, Matthijs, & Wenseleers, 2016a, 2016b) found that extra-pair paternity rates had been especially low among farmers and peasants. However, extra-pair paternity rates were considerably higher among the laboring classes in urban areas (+/- 6% per generation). Moreover, it was found that in the course of the 19th century, more or less parallel to the rise of extra-marital fertility, extra-pair paternity rates had increased significantly (Larmuseau, 2021; Larmuseau et al., 2016a). This proves that it was not only more difficult in this period for women in cities to meet and stay with the partner but also that a small, significant part cuckolded their partners.

### 3.4 HEALTH AND MORTALITY

During the 19th century important improvements in health took place, thanks to, amongst others, sanitary measures, improved nutrition, and advancements in medical knowledge and practice. As a result, mortality declined and life expectancy at birth and at later ages increased systematically. These changes are part of the so-called epidemiological transition: the shift from a mortality regime in which infectious diseases were the main causes of death towards a regime in which chronic, degenerative diseases, cancer and cardiovascular diseases prevail (Omran, 1971). In Belgium this transition was unfolding during the 19th and early 20th century, but far from complete by the end of the COR\*-research period. We observe for the period 1840–1920 a lot of epidemic outbreaks — e.g., cholera, measles and smallpox (Donrovich, Puschmann, & Matthijs, 2018; Donrovich, Puschmann, Matthijs, & Neyrinck, 2013) — and although in Belgium life expectancy at birth increased from 40 years in 1840 to 50 years by the start of World War I, the improvement was far from linear (Devos, 2006; Eggerickx, Sanderson, & Vandeschrick, 2020).

The research on health and mortality on the basis of the Antwerp COR\*-database has been concentrated on three topics: the clustering and intergenerational transfer of infant mortality, the impact of early life conditions on later life mortality (i.e., post-reproductive life expectancy), and differences in mortality risks between migrants and natives, the so-called healthy migrant effect.

In both contemporary and historical societies, deaths among infants are not spread randomly over the population. In high mortality regimes, certain families pay a high infant death toll, while others face only a limited number of deaths. This phenomenon has become known as infant mortality clustering (Das Gupta, 1990) and has received ample attention in the field of historical demography. Mattijs Vandezande (2012) was the first to explore this phenomenon using the Antwerp COR\*-database and devoted his PhD thesis to it. He approached infant mortality clustering from an individual (micro), family (meso) and population (macro) perspective. He found that high infant mortality risks were transferred from one generation to the next, both through the paternal and the maternal line, although the effects found for the maternal line were stronger. As it turned out, girls who lost a considerable share of their siblings in infancy had up to twice the risk of losing an infant as mothers, compared to mothers who had no siblings dying in infancy (Vandezande & Matthijs, 2013). Follow-up research by Robyn Donrovich et al. (2018) confirmed that women whose mothers lost three or more infants had a 77% higher risk of experiencing infant deaths among their own offspring, compared to women whose mothers lost no infants. Comparable results were found elsewhere in Europe (Quaranta et al., 2017).

Measuring infant mortality clustering in families and its intergenerational transfer is one thing, explaining it is another. Both nature, "faulty genes", and nurture explanations, "faulty parents", as well as a combination of both have been proposed (Edvinsson & Janssens, 2012; Janssens, Messelink, & Need, 2010). The research by Mattijs Vandezande (2012) pointed especially in the direction of the socialization hypothesis, stating that both health-threatening as well as health-promoting attitudes and behaviors of parents are "transmitted" from one generation to the other. Robyn Donrovich et al. (2018), by contrast, linked their results to life history theory, as they observed that mothers with the highest risk of infant mortality were more likely to bear their children at a young age, out-of-wedlock and raise them without a partner. Their limited life experience and lack of a helping hand and resources created a risky environment for their children. Offspring that survived in such circumstances followed faster and riskier reproductive strategies, creating similar risk factors among their offspring (Pink, Willführ, Volland, & Puschmann, 2020).

Infant and childhood conditions were not only important for the health and survival chances of offspring, they also had an important impact on ego's later life mortality risks (Bengtsson & Lindström, 2000; Quaranta, 2013). Donrovich, Puschmann and Matthijs (2014) investigated the impact of sibling composition (i.e., sib size and sex) and birth order on later life mortality risks in the district of Antwerp. Moreover, they investigated whether geographic proximity of siblings later in life influenced mortality risks after the age of 50. The event history models showed that sibling competition in early life profoundly impacted later life outcomes. Having older brothers was found to have had a negative effect for males' as well as females' later life survival chances. Women who had no older brothers were almost always better off. Widows were, however, an exception, as for them it was beneficial to have older brothers in later life. The same applies to widowers who had older sisters. Apparently, sibling competition gave way to solidarity and mutual support for siblings of the opposite sex in times of crisis. It is interesting to see that such mutual support was not found for men and women who remained single for life. Remaining unmarried and having a lot of siblings was especially detrimental for the survival chances of women in later life.

Another branch of COR\*-research focused on differences in survival chances (above age 30+ years) between migrants and natives. Puschmann, Donrovich, Grönberg, Dekeyser, and Matthijs (2016b) found strong evidence for a so-called "healthy migrant" effect in Antwerp, as well as in the port cities of Rotterdam and Stockholm. This is explained in terms of selection effects: only the healthiest people move, and the healthiest people move over the longest distances. The living environment was also important: migrants who grew up in the countryside, had lower mortality risks compared to natives and urban migrants. However, in the later period, when the urban penalty had turned into an urban premium, the opposite was observed: rural-to-urban migrants now faced higher mortality risks. Moreover, rural-to-urban migrants in Antwerp had higher mortality risks in times when Antwerp was hit by epidemics, probably as these migrants had not been exposed to these infectious diseases in early life and consequently lacked immunity (Alter & Oris, 2005). In a follow-up paper on Rotterdam, it was tested whether the healthy migrant effect was not caused by selective return migration of the sick and elderly, the so-called salmon bias hypothesis. The results showed that the healthy migrant effect was real, and that the most mobile migrants had the lowest mortality risks (Puschmann, Donrovich, & Matthijs, 2017).

### 3.5 MIGRATION AND SOCIAL INCLUSION

Inspired by contemporary debates on migration and social inclusion, historians of migration have studied how mobile and adaptable individuals and societies were in the past (Lucassen, 2005). Apparently, migration has always been typical for human societies, and this was also the case for 19th and early 20th-century Western societies (Hoerder, 2002). As a result of mortality decline, Europe's population grew at an unprecedented speed, while at the same time, due to innovations, less labor was needed in agriculture. Consequently, a substantial and growing part of the European population left the countryside. Increasing numbers of individuals tried their luck in other continents, especially the United States. The largest population movement was from the countryside to a nearby city, where newcomers found work in the growing industry and service sector (Moch, 2003). The urban share of Europe's population increased from about 12% in 1800 to 44% in 1910 (Clark, 2013). During this period, Antwerp turned into the largest and fastest growing city of Belgium and it became one of Europe's major ports.

Sociologists and historians have debated intensively on the extent to which urban newcomers in the 19th and early 20th century became integrated in the city. Scholars of the Chicago School of Sociology (Park, 1928; Thomas & Znaniecki, 1918) and their later followers (Handlin, 1973) argued, mostly on the basis of qualitative sources, that all migrants, but especially those who originated from the countryside, faced a very hard time to get integrated in the city. Rural-to-urban migrants lacked schooling, skills and the necessary social networks which prevented them from thriving in the new urban environment. They soon became marginalized and ended up in ghettos and poor suburbs. In their struggle for survival, a significant part grabbed the bottle, committed crime or prostituted themselves. While most of the initial work was done on American cities, Chicago in particular, European researchers reached similar conclusions on cities like Rotterdam (Bouman & Bouman, 1955) or Paris (Chevalier, 1973). Later researchers (Moch, 1983; Sewell, 1985) studied the topic with more quantitative data and came to other, sometimes even opposite conclusions: they claimed that urban newcomers were the most dynamic and enterprising city dwellers, who possessed plenty of human capital and moved into well-integrated social networks. Moreover, as the majority of them moved over short distances, they could

maintain connections with the home front. Migrants did well as they were positively selected from the population of origin: they brought the right baggage with them to the city to become successful urban dwellers. Lucassen (2004) has tried to reconcile the opposite views, by arguing that the gloomy picture applied to leavers, while the successful migrants stayed and left their traces in the quantitative source material (i.e., the registries).

Against this background Paul Puschmann (2015) studied in his PhD-thesis processes of social inclusion and exclusion among internal and international migration in Antwerp, Rotterdam and Stockholm, in the period 1850–1930. Next to the Antwerp COR\*-database, the research was based on the Historical Sample of the Netherlands and the Stockholm Historical Database. Puschmann focused on various subtopics — marriage opportunities, family formation, assortative mating, career mobility and adult mortality — in order to study to what degree migrants got access to the marriage market, other social groups, reproduction, and the labor market, and to determine whether social exclusion gave rise to certain adverse health effects in later life. By comparing three different port cities, it could be established whether or not the historical context played an important role in processes of social inclusion and exclusion.

The results regarding marriage opportunities and family formation showed that migrants had less access to marriage and reproduction compared to the native population. Migrants who originated from the vicinity of the city and arrived at younger ages, had the highest chances to get married and start a family. Generally speaking, economic capital did not increase the likelihood of getting access to the marriage market. It became clear that access to marriage was easier in Antwerp and Rotterdam, compared to Stockholm (Puschmann, 2015; Puschmann et al., 2014a; Puschmann, Van den Driessche, Grönberg, Van de Putte, & Matthijs, 2014b).

The results on partner choice learn that migrants were treated as outsiders and that this was not due to a lack of economic capital or skills, but was mainly related to ethnic and cultural characteristics. Migrants who originated from the city's hinterland and had moved early on in their life, had the highest likelihood of marrying a native partner and becoming fully part of the urban mainstream. It also became clear that migrants from smaller groups had more difficulty in getting married, but if they did, they were more likely to do so with native partners, probably because the in-group options were scarce (Puschmann et al., 2016a).

In terms of labor market inclusion, migrants also faced a hard time. Male migrants were over-represented among the lower positions, and often were not able to close the gap with the native population. However, there were differences. In general, international migrants performed better than internal migrants, and often even occupied higher positions than natives, while internal migrants were found, especially in the beginning of their career, on the lower positions. However, in Rotterdam, domestic migrants closed the gap in the course of their career, in Antwerp they even overtook natives over the life course, while in Stockholm the gap was large and even widened, showing again that social inclusion in Stockholm was more difficult, especially compared to Antwerp (Puschmann, 2015).

While on average migrants had lower mortality risks in later life than natives — the healthy migrant effect — a deeper analysis made clear that migrants paid a high price for moving to the city, as their mortality advantage decreased the longer they lived in the city. Moreover, certain sub-groups actually experienced excess mortality, sometimes due to negative selection effects, but also due to heavy and unhealthy jobs.

### 3.6 SOCIAL STRATIFICATION AND SOCIAL MOBILITY

One of the main hypotheses regarding social structure and social mobility is that societies will become meritocratic as a result of modernization, i.e., industrialization, urbanization and increased geographic mobility (Treiman, 1970). Status ascription will give way to status achievement, signaling a decrease of the (direct) influence of parents and the family on the social status and mobility chances of the children, while the importance of human capital — education, training, and skills — will increase. At the same time, overall social mobility is believed to have increased. While this hypothesis has been influential in stratification research, the evidence is mixed: some studies find little evidence for a real (relative) shift in social mobility, while others confirm that processes of modernization transformed patterns of social mobility. Wiebke Schulz, Ineke Maas and Marco van Leeuwen (2015) found, for instance, that during the latter half of the 19th and the early 20th century the association between the occupations of fathers and sons decreased in the Netherlands. Other studies find mixed results.

Richard Zijdeman (2009) found no overall effect of industrialization on social mobility in the Dutch province of Zeeland. He did, however, find an increase in social mobility in areas that experienced an expansion of the transport network.

Against this background, Cornelia Vandenberghe (2020a) studied intergenerational mobility in Antwerp in the period 1830–1913, by comparing HISCAM-scores of fathers and sons. She found evidence that Antwerp became increasingly meritocratic and that social mobility increased towards the end of the period. Overall, the likelihood of sons having a higher social status than their fathers increased over time, while the elasticity of intergenerational social mobility decreased during the study period. In line with the results of Paul Puschmann (2015), she found that migrants were more upwardly mobile than natives and that literacy increased the chances of intergenerational upward mobility among males, while at the same time the literacy of their female spouses also had an effect, suggesting that education of women (as measured by literacy status) was important for the career of their husbands.

Van Bavel, Moreels, Van de Putte, and Matthijs (2011) investigated the relationship between parent's fertility control and intergenerational social mobility of children in the city of Antwerp. Following the quality-quantity trade-off and the resource dilution hypothesis, they investigated whether the application of birth control by the parents would increase the likelihood of children to climb the social ladder. They found evidence that the children of couples that did not apply fertility control, were more likely to end at the lower ranks of the social ladder. Also, they found that fertility control limited the likelihood of children to experience downward mobility, especially among the middle classes. However, fertility limitation did not play an instrumental role to accelerate upward mobility among the offspring.

Sarah Moreels (2010) worked on female careers. Together with others, she developed GENCLASS, a historical class scheme adapted for women's occupations. GENCLASS is based on both the social power of the woman and that of her husband. This was necessary as until then, researchers had almost exclusively focused on career mobility of men, due to the under-registration of female occupations. Especially among married women, often the following occupational entries are found: "housewife" or "without occupation". Moreels also applied GENCLASS in order to study the social mobility of fertile women (ages 15 through 49) in the city of Antwerp in the period 1846–1906. She found that the social status of the majority of the studied women during family formation remained stable. This result is corroborated by the finding that parity did not predict the social status of women.

The last study we discuss is the master thesis by Cornelia Vandenberghe (2020b) on female labor market participation in Antwerp and Brussels. She compared both cities in order to determine whether the demand for female labor had an impact on their labor market participation. She found that in Brussels, where there was larger demand for female laborers in the textile industry and manufacturing, female labor market participation was not notably higher than in Antwerp, where the labor market was dominated by physically demanding jobs. Female labor market participation was rather determined by supply-side factors: unmarried women were much more likely to work than married women, higher class women were also less likely to work than laboring class women, and before marriage, migrant women were more likely to work than native women, while the reverse was found during marriage.

## 4 PATHWAYS FOR FUTURE RESEARCH

### 4.1 NEW POSSIBLE TOPICS

The Antwerp COR\*-database is relatively young (first release: 2010). As the literature review has shown, quite a number of studies have been conducted based on the analysis of the database over the past years, but this is only a fraction of what is possible. There are plenty of options for further research in all domains outlined above, but also beyond, for instance with regard to social and geographical mobility, historical family dynamics, and segregation versus inclusion. One could look into settlement patterns of migrants and natives. Did migrants mingle with natives in the housing market or did they live in segregated city parts? What were the driving forces of settlement patterns? Were individuals with port-related occupations, such as dock workers and sailors, mostly found in the neighborhoods of the port? Did individuals move into richer neighborhoods if they climbed up the social ladder? And did residential patterns influence health and survival chances? These and related questions can be studied using GIS techniques.

So far no studies on the database have looked into remarriages. Usually men remarry more often and at a faster rate (Matthijs, 2003b), but was this also the case in a port city like Antwerp with an excess number of young males, resulting from strong male in-migration? Did the skewed sex ratio translate into a stronger bargaining power of women in the marriage market and if so, how did that manifest itself? Did women remain less often single than elsewhere? Did they more often marry upward? Were women with illegitimate children more likely to remarry? And did potential female advantages disappear when in the beginning of the 20th century the sex balance became more equal? Since the context — skewed sex-ratio's — is essential in this type of exploration, a comparative IDS-project including multiple cities such as Antwerp, Rotterdam and Stockholm (Puschmann, 2015) with varying sex-ratio's seems more promising than an individual case study.

Sex-differences in mortality are another topic to further explore. Isabelle Devos (2000) showed that towards the end of the 19th and the early 20th century certain groups of girls experienced excess mortality, notwithstanding the fact that females enjoy biologically pre-determined survival advantages for all age groups. Patterns of excess female mortality have also been observed for other European countries in the same time period or even later (Weigl, 2016). Various hypotheses have been put forward to explain these excess mortality rates, ranging from females having less access to food, vaccinations and medical care to particular gender-based working and living conditions. On the basis of the COR\*-database, it can be explored which females were at an increased risk of mortality and which factors were associated with it. One could look, for example, at household composition, birth order, social status, age differences of the parents, literacy status of the parents, and living conditions.

Other topics that could be explored using the COR\*-database are (changes in) naming practices (i.e., after whom were children named), residential patterns of mothers with illegitimate children (e.g., did they cluster in certain parts of the city?), and legitimizations of illegitimate children (which children had a higher likelihood of being recognized?). Also one could dig into the life courses of (known) fathers of illegitimate children. How did their life course differ from the life course of fathers who had exclusively legitimate children? Promising are studies that will look into the influence of kin inside and outside the household on fertility and survival of offspring (e.g., grandmother hypothesis; Watkins, 2021), migration and mobility.

## 4.2 EXTENSIONS OF THE DATABASE

There are two ways of extending the database: (1) by plugging in data from other source types, and (2) by increasing the scope of the current source types (e.g., population registers and vital registration records), either through time or through space.

### 4.2.1 PLUGGING IN DATA FROM OTHER SOURCE TYPES

It can be of great added value, next to the existing demographic and socio-economic data from the population registers and vital registration, to include data from other sources. Currently the COR\*-database is already being enriched with genetic markers in the context of a postdoc project by Sofie Claerhout, funded by Research Foundation Flanders.<sup>3</sup> More specifically, currently-living COR\*-males in Flanders and the Netherlands are being sequenced on the non-recombining part of the Y-chromosome. The DNA-data that will be obtained in this way is being linked through the family tree of the sampled individuals to their ancestors in the Antwerp COR\*-database. This will shed new light on long-debated nature-nurture issues relating to surnames, family formation, migration and mortality.

Likewise, projects by third parties can lead to extensions of the COR\*-database. The IMMIBEL project "Outcast or Embraced? Clusters of Foreign Immigrants in Belgium, c. 1840–1890" led to the creation of a database with the surviving index cards of immigrants from the Public Safety Office (Sécurité Publique) as well as a database with all mariners (both Belgian and foreigners) based on the Seamen's registry of voyages.<sup>4</sup> By plugging the data on COR\*-persons and their partners into the COR\*-database, new analyses can be performed related to, amongst others, work and mobility patterns of immigrants.

The citizen-science project S.O.S Antwerp "Sociale ongelijkheid in sterfte" [Social Inequality in Death], aims to create a database of the causes of death of all individuals that died between 1820 and 1946 in

3 <https://researchportal.be/en/project/genetics-behind-historical-socio-demography-flanders-revealed-human-y-chromosome>

4 [https://www.belspo.be/belspo/brain-be/projects/IMMIBEL\\_en.pdf](https://www.belspo.be/belspo/brain-be/projects/IMMIBEL_en.pdf)

Antwerp.<sup>5</sup> If the cause of death of COR\*-individuals are linked to the Antwerp COR\*-database, this will open up research possibilities that might shed new light on the clustering and intergenerational transfer of infant and child mortality as well as the healthy migrant effect. One can investigate whether clusters of infant and child deaths can be attributed to the same causes of death, and what causes lower death rates among migrants. This can increase our insights into the drivers of mortality dynamics and differences.

New projects can be launched to collect other materials on COR\*-persons. Promising sources are conscription records, notarial deeds (e.g., prenuptial agreements, wills and property transfers), tax and rent registers. Research opportunities are, for instance, the impact of heights on life course trajectories, health and mortality and the relation of income and wealth levels with e.g., partner selection, the timing of marriage, patterns of career mobility, and health and well-being.

#### 4.2.2 VERTICAL AND HORIZONTAL EXTENSIONS

Currently, a limitation of the database is its relatively small size. Although there are observations on more than 30,000 individuals, the whole is limited in time and space. Life course data are censored by the end of the observation period (1920) and the area of observation (the district of Antwerp). For 5,971 research persons we only observe a single demographic event in the population registers and we can only link 2,459 death certificates to the 7,222 birth certificates in the database. This means that for only 35% of the individuals for which we have a birth certificate, a complete life course can be studied. That is less than 10% of all individuals in the database. For many research questions this is no problem, as event history techniques can deal well with incomplete data. However, for research questions that require a long time window the numbers soon get smaller. This is especially true if we want to include multiple generations. Robyn Donrovich et al. (2018) ended up with only 1,445 infants in their study on the intergenerational transfer of infant mortality.

Extending the data to later periods would be most interesting, but is hampered due to privacy laws. Going back in time would be possible, but it would require also a change of source type, using parish records, since population registers started not before 1846 and the civil registration not before the 1790's. While this would not lead to a similar information density as from 1846 onwards, the addition of baptism, marriage and burial records, could lead to a valuable extension of multiple generations of COR\*-individuals. This will open up new venues for research on intergenerational issues, related to marriage, fertility and mortality, but also on, for instance, naming practices, as individuals were often named after godfathers and -mothers.

Next to vertical extension in terms of time, horizontal extensions in terms of space are possible. In principle the database could be extended over the rest of Flanders or even the whole of Belgium. While this seemed until recently a "mission impossible", given the amount of time and resources it would require, new opportunities have arisen thanks to large-scale citizen science projects, as well as advances in handwritten text recognition, as successful examples from Norway and Spain have shown (e.g., Pedersen et al., 2022; Pujades-Mora et al., 2022; Thorvaldsen et al., 2015). In principle it would be possible to launch a platform and to call on volunteers — genealogists, students, retirees — to gather all entries from COR\*-persons from the population registers and vital registration records from all Belgian municipalities, provided the data collection, entry and processing are guided by data and record linkage specialists. In practice, important steps have been taken already, as the marriage certificates of Flemish Brabant and Brussels of the 19th and early 20th century have been collected in the context of the crowdsourcing project DEMOGEN, coordinated by the Belgian State Archive and KU Leuven (Matthijs, Put, & Trio, 2019). An even larger data collection has taken place for the province of West Flanders, also in the context of a citizen-science project (Aelvoet, Matsuo, Matthijs, & Buyst, 2016). By adding data on COR\*-persons from these datasets, we will be able to add a large number of new research persons, but also to complete missing life course information of existing research persons who moved from or to Antwerp at some point in time during their life.

In sum, if we overlook all the possibilities for new research on the existing database and the opportunities that might arise from future extension of the database, it seems that we are only at the beginning of a long and exciting research journey through the demographic past of Flanders and Belgium.

5 <https://sosantwerpen.be/>

## REFERENCES

- Aelvoet, A., Matsuo, H., Matthijs, K., & Buyst, E. (2016). *Linking individual historical demographic data: Province of West Flanders, Belgium*. (CeSO Working Paper, WOG/HD/2016-1). Leuven: CeSO. Retrieved from <https://lirias.kuleuven.be/retrieve/610466>
- Alter, G. (1988). *Family and the female life course. The women of Verviers, Belgium, 1849–1880*. Madison: The University of Wisconsin Press.
- Alter, G. (2020). The evolution of models in historical demography. *Journal of Interdisciplinary History*, 50(3), 325–362. doi: [10.1162/jinh\\_a\\_01445](https://doi.org/10.1162/jinh_a_01445)
- Alter, G., & Mandemakers, K. (2014). The Intermediate Data Structure (IDS) for longitudinal historical microdata, version 4. *Historical Life Course Studies*, 1, 1–26. doi: [10.51964/hlcs9290](https://doi.org/10.51964/hlcs9290)
- Alter, G., & Oris, M. (2005). Childhood conditions, migration, and mortality: Migrants and natives in 19th-century cities. *Social Biology*, 52(3–4), 178–191. doi: [10.1080/19485565.2005.9989108](https://doi.org/10.1080/19485565.2005.9989108)
- Becker, G. S. (1960). An economic analysis of fertility. In G. Becker (Ed.), *Demographic and economic change in developed countries* (pp. 209–231). Princeton: Princeton University Press.
- Bengtsson, T., & Dribe, M. (2021). The long road to health and prosperity, Southern Sweden, 1765–2015. Research contributions from the Scanian Economic-Demographic Database (SEDD). *Historical Life Course Studies*, 11, 74–96. doi: [10.51964/hlcs10941](https://doi.org/10.51964/hlcs10941)
- Bengtsson, T., & Lindström, M. (2000). Childhood misery and disease in later life: The effects on mortality in old age of hazards experienced in early life, Southern Sweden, 1760–1894. *Population Studies*, 54(3), 263–277. doi: [10.1080/713779096](https://doi.org/10.1080/713779096)
- Bouman, P. J., & Bouman, W. H. (1955). *De groei van de grote werkstad: Een studie over de bevolking van Rotterdam*. Assen: Van Gorcum.
- Bourdieu, J., Kesztenbaum, L., Postel-Vinay, G. (2014). *L'enquête TRA, histoire d'un outil, outil pour l'histoire*. Paris, INED: Classiques de l'économie et de la population.
- Caldwell, J. C. (1976). Toward a restatement of demographic transition theory. *Population and Development Review*, 2(3–4), 321–366. doi: [10.2307/1971615](https://doi.org/10.2307/1971615)
- Caldwell, J. C. (2005). On net intergenerational wealth flows: An update. *Population and Development Review*, 31(4), 721–740. doi: [10.1111/j.1728-4457.2005.00095.x](https://doi.org/10.1111/j.1728-4457.2005.00095.x)
- Caron, M., Neyrinck, W., Dillon, L. Y., Matthijs, K. (2017). Sibling influence on access to marriage: A comparative study of Québec and Flanders, 1842–1912. *Journal of Family History: Studies in Family, Kinship and Demography*, 42(3), 271–307. doi: [10.1177/0363199017713540](https://doi.org/10.1177/0363199017713540)
- Chevalier, L. (1973). *Laboring classes and dangerous classes in Paris during the first half of the nineteenth century*. New York: Fertig.
- Clark, P. (2013). Introduction. In P. Clark (Ed.), *The Oxford handbook of cities in world history* (pp. 1–30). Oxford: Oxford University Press.
- Coale, A. J., (1973). 'The demographic transition reconsidered'. In International Union for the Scientific Study of Population (IUSSP) (Ed.), *Proceedings of the International Population Conference 1973* (Vol. 1) (pp. 53–73). Liège: Editions Ordina.
- Coale, A. J., & Watkins, S. C. (1986). *The decline of fertility in Europe. The revised proceedings of a conference on the Princeton European Fertility Project*. New Jersey: Princeton University Press.
- Centre of Sociological Research, research group Family and Population Studies, KU Leuven. (2010). COR\*-database. Data Set Release 2010.08.
- Das Gupta, M. (1990). Death clustering, mothers' education and the determinants of child mortality in rural Punjab, India. *Population Studies*, 44(3), 489–505. doi: [10.1080/0032472031000144866](https://doi.org/10.1080/0032472031000144866)
- Devos, I. (2000). Te jong om te sterven. De levenskansen van meisjes in België omstreeks 1900. *Tijdschrift voor Sociale Geschiedenis*, 26(1), 55–75.
- Devos, I. (2006). *Allemaal beestjes. Mortaliteit en morbiditeit in Vlaanderen, 18de–20ste eeuw*. Gent: Academia Press.
- Donrovich, R., Puschmann, P., & Matthijs, K. (2014). Rivalry, solidarity, and longevity among siblings: A life course approach to the impact of sibship composition and birth order on later life mortality risk, Antwerp (1846–1920). *Demographic Research*, 31(38), 1167–1198. doi: [10.4054/DemRes.2014.31.38](https://doi.org/10.4054/DemRes.2014.31.38)
- Donrovich, R., Puschmann, P., & Matthijs, K. (2018). Mortality clustering in the family. Fast life history trajectories and the intergenerational transfer of infant death in late 19th- and early 20th-century Antwerp, Belgium. *Historical Life Course Studies*, 7, 47–68. doi: [10.51964/hlcs9285](https://doi.org/10.51964/hlcs9285)

- Donrovich, R. N., Puschmann, P., Matthijs, K., & Neyrinck, W. (2013). Broers en zussen: Rivalen voor het leven? Het effect van het hebben van broers en zussen en van geboorterang op de sterftkans van vijftigplussers (arrondissement Antwerpen, 1846–1910). In K. Matthijs, P. Puschmann, A. Janssens & H. Bras (Eds.), *Gender in/en historische demografie. Jaarboek historische demografie 2013* (pp. 103–130). Leuven: Acco.
- Dribe, M., & Quaranta, L. (2020). The Scanian Economic-Demographic Database (SEDD). *Historical Life Course Studies*, 9, 158–172. doi: [10.51964/hlcs9302](https://doi.org/10.51964/hlcs9302)
- Edvinsson, S., & Janssens, A. (2012). Clustering of deaths in families: Infant and child mortality in historical perspective. *Biodemography and Social Biology*, 58(2), 75–86. doi: [10.1080/19485565.2012.738575](https://doi.org/10.1080/19485565.2012.738575)
- Eggerickx, T., Sanderson, J.-P., & Vandeschrick, C. (2020). Mortality in Belgium from the nineteenth century to today. Variations according to age, sex, and social and spatial contexts. *Revue Quetelet* 8(2), 7–59. doi: [10.14428/rqj2020.08.02.01](https://doi.org/10.14428/rqj2020.08.02.01)
- Engberg, E., & Edvinsson, S. (2020). A database for the future. Major contributions from thirty-seven years of database development and research at the Demographic DataBase. *Historical Life Course Studies*, 9, 173–196. doi: [10.51964/hlcs9305](https://doi.org/10.51964/hlcs9305)
- Engelen, T. (1997). The fertility decline in the Dutch province of Limburg, 1880–1960: On understanding historical actors in a constrained environment. *The History of the Family*, 2(4), 405–424. doi: [10.1016/S1081-602X\(97\)90023-3](https://doi.org/10.1016/S1081-602X(97)90023-3)
- Fleury, M., & Henry, L. (1956). Des registres paroissiaux à l'histoire de la population: Manuel de dépouillement et d'exploitation de l'état-civil ancien. *Population*, 11(1), 142–144. Retrieved from [https://www.persee.fr/doc/pop\\_0032-4663\\_1956\\_num\\_11\\_1\\_4739](https://www.persee.fr/doc/pop_0032-4663_1956_num_11_1_4739)
- Gutmann, M., & Van de Walle, E. (1978). New sources for social and demographic history: The Belgian population registers. *Social Science History*, 2(2), 121–143. doi: [10.1017/S0145553200022367](https://doi.org/10.1017/S0145553200022367)
- Haines, M. (1998). The relationship between infant and child mortality and fertility: Some historical and contemporary evidence for the United States. In M. Montgomery & B. Cohen (Eds.), *From death to birth. Mortality decline and reproductive change* (pp. 227–253). Washington: National Academy Press. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK233807/#top>
- Hajnal, J. (1965). European marriage patterns in perspective. In D. V. Glass & D. E. C. Eversley (Eds.). *Population in history. Essays in historical demography* (pp. 101–143). New Brunswick (U.S.A.): Aldine Transaction.
- Hajnal, J. (1982). Two kinds of preindustrial household formation system. *Population and Development Review*, 8(3), 449–494. doi: [10.2307/1972376](https://doi.org/10.2307/1972376)
- Handlin, O. (1973). *The epic story of the great migrations that made the American people*. Boston: Little Brown.
- Hoerder, D. (2002). *Cultures in contact: World migrations in the second millennium*. Durham & London: Duke University Press.
- Janssens, A., Messelink, M., & Need, A. (2010). Faulty genes or faulty parents? Gender, family and survival in early and late childhood in the Netherlands, 1860–1900. *The History of the Family*, 15(1), 91–108. doi: [10.1016/j.hisfam.2010.01.005](https://doi.org/10.1016/j.hisfam.2010.01.005)
- Jenkinson, S., Anguita, F., Paiva, D., Matsuo, H., & Matthijs, K. (2020). The 2020 IDS-release of the Antwerp COR\*-database. Evaluation, development and transformation of a pre-existing database. *Historical Life Course Studies*, 9, 197–217. doi: [10.51964/hlcs9301](https://doi.org/10.51964/hlcs9301)
- Kok, J. (2005). Passion, reason, and human weakness. The European marriage pattern and the control of adolescent sexuality. In T. Engelen & A. Wolf (Eds.), *Marriage and the family in Eurasia. Perspectives on the Hajnal hypothesis* (pp. 343–367). Amsterdam: Aksant.
- Kok, J., Bras, H., & Rotering, P. (2016). Courtship and bridal pregnancy in the Netherlands, 1870–1950. *Annales de Démographie Historique*, 132(2), 165–192. doi: [10.3917/adh.132.0165](https://doi.org/10.3917/adh.132.0165)
- Kruithof, J. (1964). De demografische ontwikkeling in de XIXde eeuw. In Genootschap voor Antwerpse Geschiedenis (Ed.), *Bouwstoffen voor de geschiedenis van Antwerpen in de XIXde eeuw* (pp. 508–543). Antwerpen: Algemene Drukkerijen LLOYD Anversois.
- Larmuseau, M. H. D. (2021). De mateloze interesse in de negentiende eeuw. In J. Van Bavel, D. De Coninck, P. Puschmann & B. Van De Putte (Eds.), *Neurotic doubt and sacred curiosity. Essays in honour of Koen Matthijs/Neurotische twijfelzucht en heilige verwondering. Huldeboek voor Koen Matthijs* (pp. 141–149). Leuven: Centre for Sociological Research, Faculty of Social Sciences, KU Leuven. Retrieved from [https://www.researchgate.net/publication/355318848\\_Neurotic\\_doubt\\_and\\_sacred\\_curiosity\\_Essays\\_in\\_honour\\_of\\_Koen\\_Matthijs](https://www.researchgate.net/publication/355318848_Neurotic_doubt_and_sacred_curiosity_Essays_in_honour_of_Koen_Matthijs)

- Larmuseau, M. H. D., Claerhout, S., Gruyters, L., Nivelles, K., Vandenbosch, M., Peeters, A., van den Berg, P., Wenseleers, T., & Decorte, R. (2017). Genetic-genealogy approach reveals low rate of extra-pair paternity in historical Dutch populations. *American Journal of Human Biology*, 29(6). doi: [10.1002/ajhb.23046](https://doi.org/10.1002/ajhb.23046)
- Larmuseau, M. H. D., Matthijs, K., & Wenseleers, T. (2016a). Long-term trends in human extra-pair paternity: Increased infidelity or adaptive strategy? A reply to Harris. *Trends in Ecology & Evolution*, 31(9), 663–665. doi: [10.1016/j.tree.2016.06.012](https://doi.org/10.1016/j.tree.2016.06.012)
- Larmuseau, M. H. D., Matthijs, K., & Wenseleers, T. (2016b). Cuckolded fathers rare in human populations. *Trends in Ecology & Evolution*, 31(5), 327–329. doi: [10.1016/j.tree.2016.03.004](https://doi.org/10.1016/j.tree.2016.03.004)
- Larmuseau, M. H. D., Vanoverbeke, J., Van Geystelen, A., Defraene, G., Vanderheyden, N., Matthys, K., Wenseleers, T., & Decorte, R. (2013). Low historical rates of cuckoldry in a Western European human population traced by Y-chromosome and genealogical data. *Proceedings of the Royal Society B — Biological Sciences*, 280(1772). doi: [10.1098/rspb.2013.2400](https://doi.org/10.1098/rspb.2013.2400)
- Laslett, P. (1980). The bastardy-prone sub-society. In P. Laslett, K. Oosterveen & R. Smith (Eds.), *Bastardy and its comparative history: Studies in the history of illegitimacy and marital nonconformism in Britain, France, Germany, Sweden, North America, Jamaica and Japan* (pp. 217–246). London: Edward Arnold.
- Lesthaeghe, R. J. (2015). *The decline of Belgian fertility, 1800–1970*. Princeton: Princeton University Press. doi: [10.1515/9781400870035](https://doi.org/10.1515/9781400870035)
- Lucassen, L. (2004). De selectiviteit van blijvers. Een reconstructie van de sociale positie van Duitse migranten in Rotterdam (1870–1885). *Tijdschrift voor sociale en economische geschiedenis*, 1(2), 92–115. doi: [10.18352/tseg.826](https://doi.org/10.18352/tseg.826)
- Lucassen, L. (2005). *The immigrant threat: The integration of old and new migrants in Western Europe since 1850*. Urbana & Chicago: University of Illinois Press.
- Lynch, K. A. (1991). The European marriage pattern in the cities: Variations on a theme by Hajnal. *Journal of Family History*, 16(1), 79–96. doi: [10.1177/036319909101600106](https://doi.org/10.1177/036319909101600106)
- Mandemakers, K., & Kok, J. (2020). Dutch Lives. The Historical Sample of the Netherlands (1987–): Development and research. *Historical Life Course Studies*, 9, 69–113. doi: [10.51964/hlcs9298](https://doi.org/10.51964/hlcs9298)
- Matsuo, H., & Matthijs, K. (2016). Women's age at last childbearing and its determinants: A case study of Antwerp (Flanders, Belgium) on the basis of the multi-source COR\* historical sample. *The History of the Family*, 21(2), 175–194. doi: [10.1080/1081602X.2015.1090326](https://doi.org/10.1080/1081602X.2015.1090326)
- Matsuo, H., & Matthijs, K. (2018). The role of secularisation on marriage and conception seasonality patterns: A study of Antwerp (Flanders, Belgium) in the second half of the 19th and the early 20th century. *Journal of Family History*, 43(4), 335–356. doi: [10.1177/0363199018798745](https://doi.org/10.1177/0363199018798745)
- Matthijs, K. (2001). *De mateloze negentiende eeuw: Bevolking, huwelijk, gezin en sociale verandering*. Leuven: Universitaire Pers Leuven.
- Matthijs, K. (2002). Mimetic appetite for marriage in nineteenth-century Flanders: Gender disadvantage as an incentive for social change. *Journal of Family History*, 27(2), 101–127. doi: [10.1177/036319900202700203](https://doi.org/10.1177/036319900202700203)
- Matthijs, K. (2003a). Demographic and sociological indicators of privatisation of marriage in the 19th century in Flanders. *European Journal of Population/Revue européenne de Démographie*, 19(4), 375–412. doi: [10.1023/A:1026361006200](https://doi.org/10.1023/A:1026361006200)
- Matthijs, K. (2003b). Frequency, timing and intensity of remarriage in 19th century Flanders. *The History of the Family*, 8(1), 135–162. doi: [10.1016/S1081-602X\(03\)00010-1](https://doi.org/10.1016/S1081-602X(03)00010-1)
- Matthijs, K., & Moreels, S. (2010). The Antwerp COR\*-database: A unique Flemish source for historical-demographic research. *The History of the Family*, 15(1), 109–115. doi: [10.1016/j.hisfam.2010.01.002](https://doi.org/10.1016/j.hisfam.2010.01.002)
- Matthijs, K., Put, E., & Trio, P. (2019). Demogen: Crowdsourcing, genealogie en historisch-demografisch onderzoek in het Rijksarchief. *Science Connection*, 60, 46–48. Retrieved from [https://kuleuven.limo.libis.be/discovery/fulldisplay?docid=lirias2815380&context=SearchWebhook&vid=32KUL\\_KUL:Lirias&search\\_scope=lirias\\_profile&tab=LIRIAS&adaptor=SearchWebhook&lang=en](https://kuleuven.limo.libis.be/discovery/fulldisplay?docid=lirias2815380&context=SearchWebhook&vid=32KUL_KUL:Lirias&search_scope=lirias_profile&tab=LIRIAS&adaptor=SearchWebhook&lang=en)
- Moch, L. P. (1983). *Paths to the city: Regional migration in nineteenth-century France*. Beverly Hills: Sage Publications.
- Moch, L. P. (2003). *Moving Europeans: Migration in Western Europe since 1650*. Bloomington & Indianapolis: Indiana University Press.
- Mönkediek, B., Kok, J., & Mandemakers, K. (2016). The impact of family setting and local opportunities on leaving home and migration destinations of rural youth, the Netherlands 1860–1940. *Historical Life Course Studies*, 3, 43–65. doi: [10.51964/hlcs9353](https://doi.org/10.51964/hlcs9353)

- Moreels, S. (2010). Classifying historical occupational titles of women in a social class scheme. The career mobility of fertile women living in Antwerp during 1846–1906. *Belgisch Tijdschrift voor Nieuwste Geschiedenis*, 1–2, 143–175. Retrieved from <https://www.journalbelgianhistory.be/nl/journal/belgisch-tijdschrift-voor-nieuwste-geschiedenis-2010-1-2/classifying-historical-occupational>
- Moreels, S., & Matthijs, K. (2011). Marrying in the city in times of rapid urbanization. *Journal of Family History*, 36(1), 72–92. doi: [10.1177/0363199010390615](https://doi.org/10.1177/0363199010390615)
- Moreels, S., & Vandezande, M. (2012). Migration and reproduction in transitional times. Stopping behaviour of immigrants and natives in the Belgian city of Antwerp (1810–1925). *Historical Social Research/Historische Sozialforschung*, 37(3), 321–350. doi: [10.12759/hsr.37.2012.3.321-350](https://doi.org/10.12759/hsr.37.2012.3.321-350)
- Neven, M. (2003). Individus et familles: Les dynamiques d'une société rural. Le Pays de Herve dans la seconde moitié du XIXe siècle. Geneva: Droz. doi: [10.4000/books.pulg.5432](https://doi.org/10.4000/books.pulg.5432)
- Omran, A. R. (1971). The epidemiologic transition: A theory of the epidemiology of population change. *Milbank Memorial Fund Quarterly*, 49(4), 509–538. doi: [10.1111/j.1468-0009.2005.00398.x](https://doi.org/10.1111/j.1468-0009.2005.00398.x)
- Oris, M. (1996). Fertility and migration in the heart of the industrial revolution. *The History of the Family*, 1(2), 169–182. doi: [10.1016/s1081-602x\(96\)90005-6](https://doi.org/10.1016/s1081-602x(96)90005-6)
- Oris, M., & Ritschard, G. (2014). Sequence analysis and transition to adulthood: An exploration of the access to reproduction in nineteenth-century East Belgium. In P. Blanchard, F. Bühlmann & J.-A. Gauthier (Eds.), *Advances in sequence analysis: Theory, method, applications* (Vol. 2) (pp. 151–167). Cham: Springer. doi: [10.1007/978-3-319-04969-4\\_8](https://doi.org/10.1007/978-3-319-04969-4_8)
- Park, R. E. (1928). Human migration and the marginal man. *The American Journal of Sociology*, 33(6), 881–893. Retrieved from <http://www.jstor.org/stable/2765982>
- Pedersen, B.-R., Holsbø, E., Andersen, T., Shvetsov, N., Ravn, J., Sommerseth, H. L., & Bongo, L. A. (2022). Lessons learned developing and using a machine learning model to automatically transcribe 2.3 million handwritten occupation codes. *Historical Life Course Studies*, 12, 1–17. doi: [10.51964/hlcs11331](https://doi.org/10.51964/hlcs11331)
- Pink, K. E., Willführ, K. P., Volland, E., & Puschmann, P. (2020). Effects of individual mortality experience on out-of-wedlock fertility in eighteenth- and nineteenth-century Krummhörn, Germany. *Human Nature*, 31, 141–154. doi: [10.1007/s12110-020-09368-3](https://doi.org/10.1007/s12110-020-09368-3)
- Pujadas-Mora, J. M., Fornés, A., Ramos Terrades, O., Lladós, J., Chen, J., Valls-Fígols, M., & Cabré, A. (2022). The Barcelona Historical Marriage Database and the Baix Llobregat Demographic Database. From algorithms for handwriting recognition to individual-level demographic and socioeconomic data. *Historical Life Course Studies*, 12, 78–111. doi: [10.51964/hlcs11971](https://doi.org/10.51964/hlcs11971)
- Puschmann, P. (2015). *Social inclusion and exclusion of urban in-migrants in Northwestern European port cities. Antwerp, Rotterdam & Stockholm ca. 1850–1930*. (PhD thesis, KU Leuven). Retrieved from [https://www.researchgate.net/publication/287106072\\_Social\\_Inclusion\\_and\\_Exclusion\\_of\\_Urban\\_In-Migrants\\_in\\_Northwestern\\_European\\_Port\\_Cities\\_Antwerp\\_Rotterdam\\_Stockholm\\_ca\\_1850-1930](https://www.researchgate.net/publication/287106072_Social_Inclusion_and_Exclusion_of_Urban_In-Migrants_in_Northwestern_European_Port_Cities_Antwerp_Rotterdam_Stockholm_ca_1850-1930)
- Puschmann, P. (2020a). Introduction. In P. Puschmann (Ed.), *A cultural history of marriage in the age of empires* (pp. 1–16). London & New York: Bloomsbury Academic.
- Puschmann, P. (2020b). Love, sex and sexuality: Balancing economic considerations, sociocultural expectations, and personal desires. In P. Puschmann (Ed.), *A cultural history of marriage in the age of empires* (pp. 117–138). London & New York: Bloomsbury Academic.
- Puschmann, P., Donrovich, R., Grönberg, P.-O., Dekeyser, G., & Matthijs, K. (2016b). Disfavored in life, favored in death? Later-life mortality differences (ages 30+) between migrants and natives in Antwerp, Rotterdam and Stockholm, 1850–1930. *Historical Social Research/Historische Sozialforschung*, 41(4), 257–290. doi: [10.12759/hsr.41.2016.4.257-290](https://doi.org/10.12759/hsr.41.2016.4.257-290)
- Puschmann, P., Donrovich, R., & Matthijs, K. (2017). Salmon bias or red herring? Comparing adult mortality risks between natives and internal migrants: Stayers, returnees and movers in Rotterdam, the Netherlands, 1850–1940. *Human Nature*, 28, 481–499. doi: [10.1007/s12110-017-9303-1](https://doi.org/10.1007/s12110-017-9303-1)
- Puschmann, P., Grönberg, P.-O., Schumacher, R., & Matthijs, K. (2014a). Access to marriage and reproduction among migrants in Antwerp and Stockholm. A longitudinal approach to processes of social inclusion and exclusion, 1846–1926. *The History of the Family*, 19(1), 29–52. doi: [10.1080/1081602X.2013.796889](https://doi.org/10.1080/1081602X.2013.796889)

- Puschmann, P., Van den Driessche, N., Grönberg, P.-O., Van de Putte, B., & Matthijs, K. (2014b). Migratie en sociale in- en uitsluiting in Noordwest-Europese havensteden. Partnerkeuze en huwelijk onder migranten in Antwerpen, Rotterdam en Stockholm, 1850–1930. In I. Devos, K. Matthijs & B. Van de Putte (Eds.), *Kwestbare groepen in/en historische demografie* (pp. 115–153). Leuven: Acco. doi: [10.13140/2.1.1799.3288](https://doi.org/10.13140/2.1.1799.3288)
- Puschmann, P., Van den Driessche, N., Grönberg, P.-O., Van de Putte, B., & Matthijs, K. (2015). From outsiders to insiders? Partner choice and marriage among internal migrants in Antwerp, Rotterdam & Stockholm, 1850–1930. *Historical Social Research/Historische Sozialforschung*, 40(2), 319–358. doi: [10.12759/hsr.40.2015.2.319-358](https://doi.org/10.12759/hsr.40.2015.2.319-358)
- Puschmann, P., Van den Driessche, N., Matthijs, K., & Van de Putte, B. (2012). Marginalisatie en huwelijkssluiting onder migranten. Het acculturatieproces van migranten in de havenstad Antwerpen vanuit levensloopsperspectief (1846–1920). In K. Matthijs, J. Kok & H. Bras (Eds.), *Leren van historische levenslopen: Historisch-demografisch onderzoek in Vlaanderen en Nederland* (pp. 145–180). Leuven/The Hague: Acco. Retrieved from <https://lirias.kuleuven.be/retrieve/206487>
- Puschmann, P., Van den Driessche, N., Matthijs, K., & Van de Putte, B. (2016a). Paths of acculturation and social inclusion. Migration, marriage opportunities and assortative mating by geographic origin in Antwerp, 1846–1920. *Journal of Migration History*, 2(1), 177–207. doi: [10.1163/23519924-00201007](https://doi.org/10.1163/23519924-00201007)
- Quaranta, L. (2013). Scarred for life. How conditions in early life affect socioeconomic status, reproduction and mortality in Southern Sweden, 1813–1968 (Doctoral dissertation, Lund University). Lund: Media-Tryck. Retrieved from <https://lup.lub.lu.se/search/ws/files/3026458/3737380.pdf>
- Quaranta, L., Broström, G., van Dijk, I., Donrovich, R., Edvinsson, S., Engberg, E., Mandemakers, K., Puschmann, P., Matthijs, K., & Sommerseth, H. (2017). Intergenerational transfers of infant mortality in historical contexts: A comparative study of five European populations. Paper presented at the SSHA Conference, Chicago.
- Rotering, P. P. P. (2020). *Family influences on fertility in Europe, 1850–1920* (PhD thesis, Wageningen University). doi: [10.18174/517335](https://doi.org/10.18174/517335)
- Ruggles, S. (1999). The limitations of English family reconstitution: English population history from family reconstitution 1580–1837. *Continuity and Change*, 14(1), 105–130. doi: [10.1017/s0268416099003288](https://doi.org/10.1017/s0268416099003288)
- Schulz, W., Maas, I., & van Leeuwen, M. H. D. (2015). Occupational career attainment during modernization. A study of Dutch men in 841 municipalities between 1865 and 1928. *Acta Sociologica*, 58(1), 5–24. doi: [10.1177/0001699314565795](https://doi.org/10.1177/0001699314565795)
- Schumacher, R., Matthijs, K., & Moreels, S. (2013). Migration and reproduction in an urbanizing context. Family life courses in 19th century Antwerp and Geneva. *Revue Quetelet*, 1(1), 51–72. doi: [10.14428/rqj2013.01.01.03](https://doi.org/10.14428/rqj2013.01.01.03)
- Schumacher, R., Ryczkowska, G., & Perroux, O. (2007). Unwed mothers in the city. Illegitimate fertility in 19th-century Geneva. *The History of the Family*, 12(3), 189–202. doi: [10.1016/j.hisfam.2007.10.002](https://doi.org/10.1016/j.hisfam.2007.10.002)
- Sewell, W. H. (1985). *Structure and mobility: The men and women of Marseille, 1820–1870*. Cambridge, London & New York: Cambridge University Press.
- Shorter, E. (1971). Illegitimacy, sexual revolution, and social change in modern Europe. *The Journal of Interdisciplinary History*, 2, 237–272.
- Shorter, E. (1973). Female emancipation, birth control, and fertility in European history. *The American Historical Review*, 78, 605–640.
- Spolaore, E., & Wacziarg, R. (2021). Fertility and modernity. *The Economic Journal*, 132, 796–833. doi: [10.1093/ej/ueab066](https://doi.org/10.1093/ej/ueab066)
- Szreter, S. (1996). *Fertility, class & gender in Britain, 1860–1940*. Cambridge, New York: Cambridge University Press.
- Thomas, W. I., & Znaniecki, F. (1918). *The Polish peasant in Europe and America: Monograph of an immigrant group*. Chicago: Chicago University Press.
- Thorvaldsen, G., Pujadas-Mora, J. M., Andersen, T., Eikvil, L., Lladós, J., Fornés, A., & Cabré, A. (2015). A tale of two transcriptions. Machine-assisted transcription of historical sources. *Historical Life Course Studies*, 2, 1–19. doi: [10.51964/hlcs9355](https://doi.org/10.51964/hlcs9355)
- Treiman, D. J. (1970). Industrialization and social stratification. *Social Inquiry*, 40(2), 207–234. doi: [10.1111/j.1475-682X.1970.tb01009.x](https://doi.org/10.1111/j.1475-682X.1970.tb01009.x)

- Van Baelen, H. (2007). *Constructie van een historisch–demografisch longitudinale database. Methodologie van de Demographica Flandria Selecta*. Leuven: Centrum voor Sociologisch Onderzoek.
- Van Baelen, H., & Matthijs, K. (2007). Comparing two methods to reconstruct the kin network outside the household with nineteenth-century Antwerp data. *Journal of Family History*, 32(2), 149–160. doi: [10.1177/0363199006298064](https://doi.org/10.1177/0363199006298064)
- Van Bavel, J. (2004). Diffusion effects in the European fertility transition: Historical evidence from within a Belgian town (1846–1910). *European Journal of Population*, 20, 63–85. doi: [10.1023/B:EUIP.0000014572.66520.0d](https://doi.org/10.1023/B:EUIP.0000014572.66520.0d)
- Van Bavel, J., Moreels, S., Van de Putte, B., & Matthijs, K. (2011). Family size and intergenerational social mobility during the fertility transition: Evidence of resource dilution from the city of Antwerp in nineteenth century Belgium. *Demographic Research*, 24(14), 313–344. doi: [10.4054/DemRes.2011.24.14](https://doi.org/10.4054/DemRes.2011.24.14)
- Vandenberghe, C. (2020a). *Like father, like son? Industrialization and intergenerational social mobility: evidence from Antwerp, 1830–1913* (Research paper). London: London School of Economics.
- Vandenberghe, C. (2020b). *Supply or demand? An evaluation of the determinants of Belgian female labour force participation during industrialization: Antwerp and Brussels in the early twentieth century* (Master thesis). London School of Economics, London.
- Vandezande, M. (2012). *Born to die. Death clustering and the intergenerational transmission of infant mortality, the Antwerp district, 1846–1905* (PhD thesis, KU Leuven). Retrieved from <https://lirias.kuleuven.be/retrieve/206174>
- Vandezande, M., & Matthijs, K. (2013). Inherited dimensions of infant mortality. Detecting signs of disproportionate mortality risks in successive generations. *The History of the Family*, 18(2), 169–186. doi: [10.1080/1081602X.2012.755131](https://doi.org/10.1080/1081602X.2012.755131)
- Vries, S. H. (2019). *Ongehuwde moeders en voorechtelijke verwekkingen in het arrondissement Antwerpen, 1812–1920* (Bachelor thesis, Radboud University Nijmegen). Retrieved from [https://theses.uibn.ru.nl/bitstream/handle/123456789/10471/Vries%2c\\_S.H.\\_1.pdf?sequence=1](https://theses.uibn.ru.nl/bitstream/handle/123456789/10471/Vries%2c_S.H._1.pdf?sequence=1)
- Watkins, A. (2021). Reevaluating the grandmother hypothesis. *History and Philosophy of the Life Sciences*, 43(103). doi: [10.1007/s40656-021-00455-x](https://doi.org/10.1007/s40656-021-00455-x)
- Weigl, A. (2016). The drifting apart of gender-specific life expectancies in Europe 1850–2010. Some introductory remarks. In M. Dinges & A. Weigl (Eds.), *Gender-specific life expectancy in Europe 1850–2010* (pp. 7–15). Stuttgart: Franz Steiner Verlag. Retrieved from [file:///ia/IISG\\$/home/MarjaK/Downloads/MedGG\\_Beiheft\\_58-1.pdf](file:///ia/IISG$/home/MarjaK/Downloads/MedGG_Beiheft_58-1.pdf)
- Westberg, A., Engberg, E., & Edvinsson, S. (2016). A unique source for innovative longitudinal research: The POPLINK Database. *Historical Life Course Studies*, 3, 20–31. doi: [10.51964/hlcs9351](https://doi.org/10.51964/hlcs9351)
- Winter, A. (2009). *Migrants and urban change. Newcomers to Antwerp, 1760–1860*. London: Pickering & Chatto.
- Zijdeman, R. L. (2009). Like my father before me: Intergenerational occupational status transfer during industrialization. *Continuity and Change*, 24(3), 455–486. doi: [10.1017/S0268416009990166](https://doi.org/10.1017/S0268416009990166)

This edited volume discusses the impact of several major databases containing historical longitudinal population data. The creation and development of these databases have greatly expanded research possibilities in history, demography, sociology, and other disciplines. The present collection includes seven contributions, on eight databases, that had a wide impact on research in various disciplines. Each database had its own unique genesis and readers are informed about how these databases have changed the course of research in historical demography and related disciplines, how settled findings were challenged or confirmed, and how innovative investigations were launched and implemented.

The volume serves as an essential resource for scholars in the field of historical life course studies, offering insights into the transformative power of these databases and their potential for future advancements in academic research.

**Sören Edvinsson** is Professor Emeritus of Historical Demography at the Centre for Demographic and Aging Research at Umeå University (CEDAR).

**Kees Mandemakers** is affiliated member of the Radboud Group for Historical Demography and Family History. He is senior fellow at the International Institute of Social History and Emeritus Professor of Large Historical Databases at the Erasmus School of History, Culture and Communication of Erasmus University Rotterdam.

**Ken R. Smith** is Distinguished Professor of Family Studies and Population Science and former Director of the famous Pedigree and Population Resource at the University of Utah, as well as Investigator at the Huntsman Cancer Institute.

**Paul Puschmann** is Assistant Professor of Economic, Social and Demographic History at Radboud University Nijmegen.

ISBN 978-94-9329-618-3



9 789493 296183 >

**Radboud Universiteit**



[www.radbouduniversitypress.nl](http://www.radbouduniversitypress.nl)