

Introduction: Intergenerational Transmissions of Infant Mortality using the Intermediate Data Structure (IDS)

By Luciana Quaranta & Hilde Leikny Sommerseth

To cite this article: 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. <http://hdl.handle.net/10622/23526343-2018-0014?locatt=view:master>

HISTORICAL LIFE COURSE STUDIES

Intergenerational transmissions of infant mortality using the Intermediate Data Structure (IDS)

VOLUME 7, SPECIAL ISSUE 2
2018

GUEST EDITORS
Luciana Quaranta
Hilde L. Sommerseth



MISSION STATEMENT

HISTORICAL LIFE COURSE STUDIES

Historical Life Course Studies is the electronic journal of the *European Historical Population Samples Network* (EHPS-Net). The journal is the primary publishing outlet for research involved in the conversion of existing European and non-European large historical demographic databases into a common format, the Intermediate Data Structure, and for studies based on these databases. The journal publishes both methodological and substantive research articles.

Methodological Articles

This section includes methodological articles that describe all forms of data handling involving large historical databases, including extensive descriptions of new or existing databases, syntax, algorithms and extraction programs. Authors are encouraged to share their syntaxes, applications and other forms of software presented in their article, if pertinent, on the EHPS-Net website.

Research articles

This section includes substantive articles reporting the results of comparative longitudinal studies that are demographic and historical in nature, and that are based on micro-data from large historical databases.

Historical Life Course Studies is a no-fee double-blind, peer-reviewed open-access journal supported by the European Science Foundation (ESF, <http://www.esf.org>), the Scientific Research Network of Historical Demography (FWO Flanders, <http://www.historicaldemography.be>) and the International Institute of Social History Amsterdam (IISH, <http://socialhistory.org/>). Manuscripts are reviewed by the editors, members of the editorial and scientific boards, and by external reviewers. All journal content is freely available on the internet at <http://www.ehps-net.eu/journal>.

Co-Editors-In-Chief:

Paul Puschmann (Radboud University & KU Leuven) & Luciana Quaranta (Lund University)
hislives@kuleuven.be

The European Science Foundation (ESF) provides a platform for its Member Organisations to advance science and explore new directions for research at the European level. Established in 1974 as an independent non-governmental organisation, the ESF currently serves 78 Member Organisations across 30 countries. EHPS-Net is an ESF Research Networking Programme.



The European Historical Population Samples Network (EHPS-net) brings together scholars to create a common format for databases containing non-aggregated information on persons, families and households. The aim is to form an integrated and joint interface between many European and non-European databases to stimulate comparative research on the micro-level.
Visit: <http://www.ehps-net.eu>.



Introduction: Intergenerational Transmissions of Infant Mortality using the Intermediate Data Structure (IDS)

Luciana Quaranta
Lund University

Hilde Leikny Sommerseth
UiT - The Arctic University of Norway

ABSTRACT

It has previously been shown that infant mortality clusters in a subset of families, a phenomenon which was observed in historical populations as well as contemporary developing countries. A transmission of death clustering across generations has also been shown in Belgium, but it is unknown whether such effects are specific to the studied context or are also found in other areas. The current article introduces a special issue devoted to analysing intergenerational transmissions of infant mortality across the maternal line in Belgium, the Netherlands, northern and southern Sweden, and Norway. Taking advantage of the Intermediate Data Structure (IDS), the five empirical studies created datasets for analysis and ran statistical models using exactly the same programs, which are also published within the special issue. These works are the first set of studies using the IDS on several databases for comparative purposes. Consistent results across the studied contexts were shown: transfers of infant mortality across the maternal line were seen in all five areas. In addition, the works have shown that there are large advantages of adopting the IDS for historical demographic research. The structure has in fact allowed researchers to conduct studies which were fully comparable, transparent and replicable.

Keywords: : Infant mortality, Intergenerational transmissions, Intermediate Data Structure, Historical demography, Longitudinal data

e-ISSN: 2352-6343
PID article: <http://hdl.handle.net/10622/23526343-2018-0014?locatt=view:master>
The article and the software can be downloaded from [here](#).

© 2018, Luciana Quaranta, Hilde Leikny Sommerseth
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

Over the past decades great advancements have been made in historical demographic research, primarily thanks to the fact that many more longitudinal databases have become available, as well as to major developments in computer software. Vast knowledge has been produced focusing on identifying and explaining the role of different factors in determining socioeconomic attainment, marriage, reproduction, migration, health and mortality. It has been shown that demographic outcomes are not only affected by individual level characteristics, but that family and contextual level factors also play a focal role.

Regarding early life mortality, we have over the past 25 years witnessed an increased attention in the literature towards the phenomenon of mortality clustering, i.e. that infant mortality tends to cluster in certain families, a pattern observed both in historical populations (e.g. Brändström, 1984; Edvinsson, Brändström, Rogers, & Broström, 2005; Janssens, Messelink, & Need, 2010; Vandezande, 2012) as well as in contemporary developing countries (Das Gupta, 1990). Said differently, while the majority of families do not experience any infant deaths, a more limited group of families experience multiple infant deaths. These findings have highlighted the importance of considering the family, instead of a single child, as the unit of analysis when examining outcomes concerning children's survival (Edvinsson et al., 2005). By taking the varying risk between families into account, we know from previous research that certain demographic characteristics had significant negative effects on the survival chances during first year of life. For instance, negative effects of shortened birth intervals due to death in infancy of an older sibling, remarriage of the mother, and experiences of stillbirths have been found in historical as well as contemporary societies (e.g. Edvinsson et al., 2005). In addition, social and cultural factors related to childcare behaviour, education and socioeconomic status (SES), as well as shared disease environment, explain part of the variation observed in risks of infant deaths across families (e.g. Brändström, 1984; Janssens et al., 2010; Reid, 2001; Scalone, Agati, Angeli, & Donno, 2017). Further, mortality clustering appears not to be limited to only one generation, in the sense that parents who lost many siblings in their infancy were more prone to experience infant deaths of their own offspring (Vandezande, 2012). In other words, it has also been demonstrated that infant mortality is transmitted along maternal as well as paternal lines. The intergenerational component gave rise to a typical nature-nurture debate, which can be summarized as 'faulty genes versus faulty parents' (Janssens et al., 2010). It remains, however, unclear to which extent intergenerational transmissions in infant mortality can be found in other regions characterized by different mortality regimes, and to which extent it can be explained by other demographic characteristics of the family.

In a recently published literature review about early-life mortality clustering in families, Van Dijk (2018) distinguished between four methodological approaches; (i) studies that focus on the influence of the death of an older sibling, (ii) mother heterogeneity, where family- and child-level characteristics are taken into account, (iii) studies explicitly focusing on mortality profiles on the family level (high- and low-risk families), (iv) studies that explore to what extent excess deaths in populations can be explained by the characteristics of children and families. This special issue focuses on approaches (i) and (iii), where maternal heterogeneity, on the threshold of high- or low-risk families, are assessed across generations.

Even though the availability of longitudinal micro-level data has allowed great advancements on research in historical demography, there are still many challenges faced in the use and scope of such type of data. Firstly, due to the complexity of changed perceptions and conceptualization over time, both spatially and relationally, researchers need advanced data management skills in order to create the datasets suitable for analysis. Moreover, the digitization and construction of historical population registers are time consuming, meaning that most databases and studies conducted on them relate to small communities, making the results of such work non-generalizable. In addition, most databases available have been developed using independent data structures, and datasets and programs are not publicly available, making the majority of published works non-replicable. Few comparative studies have been conducted, and while almost all have been based on common research questions and model definitions, they have used independent programs to create the datasets for analysis and to run the statistical models. It is therefore difficult to distinguish whether eventual differences found in the results are due to actual dissimilarities across the different studied contexts or due to differences in the data handling and model generating processes.

To try to overcome some of these challenges, the Intermediate Data Structure (IDS) was developed as a strategy aimed at standardizing the dissemination of micro-level historical demographic data (Alter & Mandemakers, 2014; Alter, Mandemakers, & Gutmann, 2009). The structure provides a common intermediate format for data from different databases, regardless of their original form. The main motivation has been to create a common format for the standardization and dissemination of data, which allows for

comparative studies across different databases, as well as to develop and exchange software for analysis. In this way the IDS is open for increased transparency, replicability and generalizability of research. From its launch in 2009, several databases have started to implement the structure, but no comparative empirical studies have been conducted to date.

The articles included in this special issue of *Historical Life Course Studies* are part of an international project which has two main aims. The first aim is to contribute to the literature on intergenerational transmissions of infant mortality, by looking at transfers across the maternal line in five different historical populations in Europe – Belgium (Donrovich, Puschmann, & Matthijs, 2018), the Netherlands (Van Dijk & Mandemakers, 2018), Norway (Sommerseth, 2018), Northern Sweden (Broström, Edvinsson, & Engberg, 2018) and Southern Sweden (Quaranta, 2018a). A comparison between the regions was also made (Quaranta et al., 2017). The five studies are based on databases that have been formatted to follow the IDS and the second aim of this project is therefore to test the capabilities of the IDS for comparative research using longitudinal historical demographic databases. These works are in fact the first set of empirical studies using the IDS on several databases for comparative purposes. These studies are all conducted using the same methods and theoretical framework, and the same programs were used to create each dataset for analysis and to run the statistical models (Quaranta, 2016; 2018b). The type of research topic considered – intergenerational transmissions – fits perfectly well with the aim of testing the IDS and presenting its advantages, given that datasets for analysis are created by constructing variables that link together three generations of individuals, following them across both time and space. The main research question for all five papers was to investigate the extent to which the level of infant mortality amongst the children of mothers in a ‘first generation’ correlated with the infant mortality level amongst the children of their daughters, the ‘second generation.’ We hypothesized that there was intergenerational transmission of mortality risk down the maternal line, such that daughters would have faced an increased risk of death amongst their own infants if their mothers had also experienced infant deaths.

In this introduction to the current special issue, we first outline issues concerning the implementation of the IDS and the different databases used. Next, we present the five regions included in the project. After that, we provide a summary of the results obtained in each of the independent articles. Then we summarize our experiences in conducting an IDS-based international collaboration and in the development of common software. Finally, we present our conclusions, focusing on the two main objectives of this project.

2 SOURCE MATERIAL AND DATA STRUCTURE

The five works included in this special issue have all used a common script for selection of cases, construction of variables and for the statistical analysis (Quaranta, 2016; 2018b). This is made possible thanks to the fact that all databases follow the IDS. Some of the databases were already transformed before the start of the project, while others were transformed for the purpose of participating in the collaboration. The IDS functions as an in-between format where longitudinal datasets are transformed into a similar structure, ready for researchers who are interested in global trends of demographic patterns, and for researchers who want to be a part of an international community where data and statistical software programs are shared.

The threshold for many historical demographers to start working on transnational comparisons is that each country's data has its own format and definitions, either based on population registers alone or family reconstitution, or a mix of registers and census enumerations. The IDS aims to remove this threshold. Undoubtedly, with a common standardization and dissemination of the data, the IDS will make comparisons across databases easier and with greater accuracy than is the case with comparative studies based on independent databases. However, what are the downsides, if any, for a less time consuming, seemingly more accurate, and more cost-efficient method? The databases need to be transformed into IDS, which requires resources and also time to learn a new data structure. Moreover, from a database theory perspective, the IDS is not the most elegant nor memory-efficient way of storing data. However, the significance of these costs fades off the more studies that are made using the data, particularly studies that are comparative or that use readily available software.

The five contributions to this special issue use data from the COR* database for Antwerp (Belgium – Matthijs & Moreels, 2010), the LINKS database for Zeeland (the Netherlands – Mandemakers & Laan, 2017), the POPUM database for Skellefteå (northern Sweden – Westberg, Engberg, & Edvinsson, 2016), the

SEDD database for Scania (southern Sweden – Bengtsson, Dribe, Quaranta, & Svensson, 2014), and the Historical Population Register (HPR) database for Troms (Norway – Thorvaldsen, Andersen, & Sommerseth, 2015). These databases contribute with a similar basket of attributes given to selected individuals for this study, where the main interest is the date of vital events, such as birth, marriage and death. Table 1 provides a summary description of the databases and sources used to construct them.

Table 1 *Description of the databases included in the project*

Name of database/ Country	Areas	Sources	Type of sources	Vital events
LINKS/ The Netherlands	Zeeland	Civil registers	Certificates	Births, marriages and deaths
POPUM/ Sweden	Skellefteå	Church books	Birth, marriage, death and migration registers Catechetical examination registers	Continuously updated information on household composition, births, marriages, deaths and migration
SEDD/ Sweden	Scania (parishes: Hög, Kävlinge, Halmstad, Sireköpinge, Kågeröd)	Ditto	Ditto	Ditto
HPR/ Norway	Troms	Church books and population censuses	Baptism, marriage and burial registers Census records	Births, marriages and deaths Decadal census registration
COR*/ Belgium	Antwerp	Population and civil registers	Population registers and birth, marriage and death certificates	Continuously updated information on household composition, migration and births, marriages and deaths

As shown in Table 1, the different databases originate from diverse sources with different initial purposes. The most obvious contrasts are the division between sources originating from the official administration's need to monitor the population movement through civil registers (the Netherlands and Belgium), and the church's need to monitor the religious demands given to its parishioners (Norway and Sweden). Obviously, with the increased expansion of state administrative tasks, the priest became an important provider of information regarding population movements within his parish, a development that is also mirrored in the increasingly more detailed reforms of the church books in both Norway and Sweden. Does this difference matter? Yes, at least in two ways. First, since our primary task is to 'reconstruct' life courses based on births, marriages and deaths, *dates* of these events are of obvious importance. While the civil registers clerics wrote certificates that documented the actual event of the birth, marriage or death, priests documented the sacred ceremony of a baptism, marriage or funeral. The IDS has a built-in solution to distinguish between these kinds of differences, by making a distinction between four different date types. *Events*, refers to a date of an event, observed at the moment of the event itself, like the birth certificate or the baptism of a child. *Reported* date, refers to the date of an event reported in a later source, such as the birth date registered upon the baptism. *Declared* date, is a date that refers to a point in time or period a certain attribute was valid. 'Married' in a census can serve as a typical example, and common for the declared dates was that the event it refers to happened at another date than the date used in the census (as this example points to). Finally, *Assigned* date refers to a date assigned by the database administrator. In the end, a common point of view is that we use the data available, with all its biases. A reasonable question would be if we gain something valuable out of the date distinction or is it just ignorable noise? Making a distinction between type of dates emphasizes the

importance of *time* between registration of the event, and the actual event itself. Hence, firstly it leads the user (researcher) to be aware of the importance of source criticism. Sources that *report* a date of an event, for example two years after the event occurred, may have other related source-critical issues such as possible under registrations, compared to sources where the registration of the event took place in connection to the event – such as date of marriage in a marriage book. It also urges the constructor of the database to reflect in his or her documentation on source criticism by asking the basic questions about the origin of dates present in the sources. The focus on different types of dates has enabled the project to test the possible effect of a time lag between birth and baptism on the distribution of mortality, as we should expect some bias in the reporting of deaths, given that the deaths of children who died before they were baptized are more likely to have been under-reported.

The databases differ not only in terms of the sources they are based on, but also in the way they are constructed. The Dutch and Norwegian databases can be defined as family reconstitution (or a variant of that), where life courses are 'reconstructed' based on marital unions, where births and deaths are added successively to the family tree. In these databases, observations on life courses are fragmentary and occur in concordance with vital events of persons or their relatives, such as their children and spouses. The two Swedish databases and the Belgian database have one clear advantage over the other databases, and that is the continuous monitoring of individuals whereabouts registered on the level of a household, in the catechetical examination register (*hüsförhörslängder*) for the Swedish databases, and population registers for the Belgian database. Each population register in Belgium was based on a ten-year census, but was also continuously updated by registering all migrations and vital events occurring between such censuses. Similarly, each catechetical register in Sweden typically covered a five-year period, and was also updated continuously.

In all studies the sample selected for analysis consisted of women with known information about the births and infant mortality/survival of their offspring and with known information about the births and infant mortality/survival of the offspring of their mothers (the 'grandmothers'). In addition, grandmothers must have given birth to at least two children, one of whom was a woman who survived to adulthood and became a mother herself. The differences in the type of source material that the different databases were based on do not only lead to different approaches for record linkage (with a greater flexibility and higher scores for the population registers), but they may also lead to challenges regarding sample selection. As has been pointed out by the Dutch group, the ideal situation for this study would have been complete observation period for grandmothers and mothers, e.g. from birth or in-migration, to death or out-migration. Instead, each individual database operates to some extent with his/her own rules for selection. Compared to the Swedish and Belgian databases, the Dutch and Norwegian databases have to rely on more fragmented observations for observing individuals during a long-time span, and this makes it more difficult to decide the selection criteria.

The databases also differ in their coverage. Those of Skellefteå, Scania, Troms and Zeeland cover the full population of the studied areas. Instead the database for Antwerp was created by following individuals whose family name started with the letters 'C-O-R'. The study areas differed in terms of size and of their population. The study sample included 13,053 infants in Skellefteå, 8,770 in Scania, 25,730 in Troms, 203,802 in Zeeland, and 1,441 in Antwerp.

3 THE REGIONS COMPARED

The special issue includes three papers from Nordic countries, two on Sweden and one on Norway. One of the papers on Sweden covers the rural area of Skellefteå in the province of Västerbotten. This area had a high population but was also large in terms of size. It was, in fact, considerably larger than most rural parishes in Sweden. The main part of the population resided in the coastal area and in river valleys. During the 19th century, infant mortality rate (IMR) was low and fertility was high. The parish was demographically recognized, with population increasing primarily because of high natural growth. For the cohorts of children considered, 1844-1901, IMR was 83 per 1,000 in the study sample. The second paper from Sweden covers five rural parishes in Scania, the southernmost province of Sweden. All five parishes are typically inland peasantry areas, and regarding demographic characteristics, quite similar to that of the Skellefteå province. For the cohorts of children considered, 1740-1968, IMR was 133 per 1,000 in the study sample. Given the length of the study period, large declines in IMR were observed across such years, with values of up

to 450 children per 1,000 in the mid-18th century and of about 50 per 1,000 in the mid-20th century. IMR fluctuated around similar values as Skellefteå if the same cohorts are considered in the two areas. The paper on Norway covers a large part of the second northernmost province, Troms. People were settled in the coastal areas, along rivers, and in the inland area all the way to the Swedish or Finish national borders. During the study period the area was characterised by low age at marriage and high fertility, and a general decrease in mortality, primarily driven by a decline in child mortality. For the cohorts of children considered, 1844-1924, IMR was 80 per 1,000 in the study sample.

The special issue also includes one paper on the Netherlands and one on Belgium. On the Netherlands, we have a paper covering the south-western coastal province of Zeeland. In comparison with Sweden and Norway, Zeeland's population experienced a much higher IMR, 164 per 1,000 in the study sample for the cohorts of children considered, 1833-1912. Across the study period, IMR halved, starting from values of about 250 per 1,000 in the early 19th-century and ending with values of about 120 per 1,000 in the early 20th-century. A bit south of Zeeland, crossing the border to Belgium, we have a paper covering the Antwerp district, which consisted of more than fifty rural municipalities and the city of Antwerp. Like the other areas, Antwerp experienced a rapid population increase during the period of study. This is the only study areas among the five considered in the special issue that comprised a major city. As in Zeeland, IMR was also high compared to the Nordic populations studied. It was 170 per 1,000 in the study sample for the cohorts considered, 1839-1915. It should be noted, though, that at the start of the study period IMR was low (around 50 per 1,000), but IMR increased substantially across the study period. Such a pattern was likely an effect of increases in population density that resulted from strong urban in-migration (Puschmann, 2015). Similar patterns of increasing IMR across the same periods were also observed in the parts of the Netherlands that neighboured Antwerp (Van den Boomen & Ekamper, 2015).

4 RESULTS OF THE STUDIES

The articles from Antwerp, Zeeland, Troms and Scania shared a common basic statistical model to begin with, and afterwards extended the analysis in various ways. The statistical analysis conducted in Skellefteå differed somewhat from the other papers, and focused not only in establishing associations but also on testing different approaches of analysis to measure intergenerational transmissions in infant mortality. Such work showed that simple models with few restrictive assumptions gave similar results as more complicated models.

All papers but one (Antwerp) found significant evidence of intergenerational continuities of infant mortality along the maternal line. Although the results were not statistically significant for Antwerp, likely due to small numbers, the patterns observed were still the same in this region as in the other four areas and statistically significant results were observed when re-categorizing the variable measuring infant mortality, as is described below. Comparing the hazards of death in infancy if the maternal grandmother had experienced infant deaths in the high-mortality regime of Zeeland (12-34% increase) with the hazards from relatively low mortality regimes found in Norway (21-28% increase) and Sweden (5-24% in Scania and 13-42% in Skellefteå), the patterns observed are quite equal, and not in the direction we expected given the higher underlying mortality rates for Zeeland.

The basic models estimated for Antwerp, Zeeland, Troms and Scania included controls for the number of births of the maternal grandmother, the child's birth order, sex and birth date, and the mother's age at birth. The results for control variables varied across the studied populations, in terms of direction, magnitude and statistical significance. For example, the risk of death in infancy was higher for boys than for girls in Scania and in Zeeland, no distinct differences were seen in Antwerp, while, surprisingly, the risk of death was lower for boys than for girls in Troms, although such differences were not statistically significant.

The works from Antwerp, Zeeland, Troms and Scania also estimated a series of extended models, focusing on different sets of variables. For Antwerp, intergenerational transmissions in infant mortality was first re-estimated by categorizing the variable into 0, 1, 2 and 3+ deaths, instead of 0, 1 2+ deaths as in the basic models. When doing so, higher and statistically significant risks of death were observed for children whose maternal grandmothers had experienced three or more deaths. The study also estimated models to dig deeper into the influence of parental factors on infant mortality risk, which showed increased risks of death in

infancy for children born to young mothers, even when controlling for socioeconomic status and birthplace, for children born to mothers who had previously worked in domestic service or in the child care sphere, and for children whose fathers were older than 40.

The extended analyses conducted for Zeeland focused on evaluating the role of survival of parents and of the socioeconomic status of the father and grandfather. Even if parental death has strong impact on the survival of infants, the results did not show any evidence that parental death plays an important role in explaining intergenerational transmission in infant mortality. Similarly, socioeconomic status of the mother and grandfather influenced the survival of infants, but socioeconomic status only played a small role in explaining mortality clustering among mothers.

The extended models conducted for Troms considered a different way of categorizing the variable measuring infant mortality in the grandmother's generation, by identifying risks groups based on the infant mortality rate rather than a pure count of infant deaths. The results showed steady increases in the risk of death in infancy according to whether the grandmothers were classified as having had 'zero', 'low', 'medium' or 'high' rates of infant mortality.

The extended models estimated for Scania considered differences in the effects of the intergenerational transmission variable by SES, time-period and gender. The effects remained consistent when controlling for SES of the father. However, when interacting such variable with SES, children whose maternal grandmother experienced 2 or more infant deaths, they still showed higher risk of dying in infancy, but the hazards declined with increasing occupational score. When making distinctions by gender, girls whose maternal grandmothers experienced infant deaths had higher risks of death in infancy, but not boys. Finally, when dividing the study into two sub-periods, significant effects of the intergenerational transmission variable were seen for children born in 1740-1865, but not for those born in 1865-1968.

The overall conclusions that can be drawn from the different models estimated across the five studies is that irrespective of geographical location, climatic and topographical variances, social, economic and cultural differences, in all areas higher risks of death in the first year of life were observed for children whose maternal grandmother had experienced the death of two or more of her offspring in the past. Such transmissions were not explained by basic demographic characteristics of the mother nor by socioeconomic status. The effects of other variables included as controls in the models, instead, varied in direction, magnitude and statistical significance across the studied areas.

5 EXPERIENCES IN CONDUCTING AN IDS-BASED COLLABORATION AND IN THE DEVELOPMENT OF COMMOM SOFTWARE

The collaboration that led to the current special issue in *Historical Life Course Studies* was initiated in one of the IDS meetings held at Lund University. The first results of the five studies were presented at the second conference of the European Society of Historical Demography, held in Leuven, Belgium in September 2016. Throughout the course of this collaboration, the work has been rather straightforward, and was primarily conducted locally by each group at their own university and through online meetings. Having a common and clear data structure to work with allowed the group to easily define variables, concepts and models.

One of the main challenges experienced within this project occurred in the initial stages of the work. When the project was initiated not all participants had fully converted their databases into IDS and most participants had not previously used IDS for research. Various problems were therefore encountered initially, which were mainly related to the fact that not all parts of the tables of each database complied fully with the requirements set by the IDS. After fixing such issues it was possible to develop common programs to create the datasets for analysis and run the statistical models, and for all participants to easily use the programs.

The programs used to create the datasets for analysis and to run the statistical models are discussed in detail in the second article of the special issue (Quaranta, 2018b). Developing such programs was a complex task. However, this complexity was primarily due to the fact that it is difficult to understand conceptually as well as to write code to create variables and datasets by following individuals longitudinally across their lives and across generations, rather than the fact that different databases were used. Writing common programs for the five datasets in the project, which could also be used by any other IDS dataset, was not significantly

more complex than writing programs to be used by one single database. Overall, thanks to the fact that all databases had adopted the same structure, the total amount of programming required across the project was substantially less than the total amount of programming that would have been required had each database used its own independent database structure. The only additional challenge faced in developing a common code to select data was the need to make the code compatible for databases constructed using solely registers of births/baptisms, marriages and deaths/burials (i.e. the databases for Norway and the Netherlands), which only contain information on vital events, as well as for databases constructed using population registers (i.e. the Belgian and Swedish databases), which in addition to vital events contain information about migration. The main factor that distinguishes the two types of databases is how periods in which individuals are under exposure are defined. This factor is very important, not only for the current project but for a variety of research questions using longitudinal methods as well.

In terms of limitations of an IDS-based approach, using exactly the same statistical models reduces the flexibility that may be adopted in each study population, and models do not necessarily fit as well in one population as in the next. For example, statistically significant effects of the number of infant deaths of the maternal grandmother were seen for Antwerp if the variable was categorized into 0, 1, 2 and 3+ deaths, but not if it was categorized into 0, 1 and 2+ deaths. In the other four areas, instead, the latter categorization gave significant results. One further limitation was that the way the study populations were selected differed across some of the areas because of differences in the structure of databases based on family reconstitution and on population/parish register data. In order to account for these issues, in a sensitivity analysis further models were estimated, making different restrictions when selecting the study sample, in terms of observation period and censoring. Such analyses showed no changes in the results in relation to the main explanatory variable. A further limitation is that starting years of study differed between the datasets, which is a selection requirement that varies by definition between the datasets. Regardless of these limitations, there were many advantages of using a common approach. The development of common programs has allowed each research group to obtain model results from their data by simply running the programs, and, in addition, the results of the research are fully comparable across the studied populations. We were able, in fact, to compare the results without running the risk that eventual differences in the patterns observed were due to differences in the data generating process or in the way models were set up and estimated.

6 CONCLUDING REMARKS

The articles included in this special issue had two main aims. The first aim was to advance the knowledge about intergenerational transmissions in infant mortality. In the five European populations studied, we observed that the likelihood that a woman's offspring died in infancy was influenced by whether any of her siblings had died in infancy. On a general level, higher infant mortality rates were seen in Antwerp and Zeeland than in Scania, Skellefteå and Troms, and infant mortality increased across the study period in Antwerp while it declined in the other four territories. Regardless of such differences in underlying infant mortality rates as well as of differences in the time periods and size of the territories considered, and their cultural and geographical characteristics, intergenerational transmissions in infant mortality was consistently found in these areas. Such transmission could not be explained by basic demographic characteristics of the mother. These results show that when studying child health and mortality it is important to consider the characteristics of the mother and even the grandmother.

The second aim of this work was to test the IDS for the first time in an international comparative project and in this way identify the advantages of using the IDS in historical demographic research. Within our collaboration we were able to create variables in an identical fashion and to make data selections and run statistical models in exactly the same way. Although there is some investment involved in transferring databases into IDS, the studies in this special issue show that by adopting this structure it is possible to develop common software to produce datasets for analyses and run statistical models, and to publish this software in Open Access format. The results of the studies are not only fully comparable, but also more transparent and replicable. By conducting fully comparative research, the conclusions reached are no longer context-specific, but instead generalizable. The IDS can therefore not only break down many of the challenges that relate to the difficulty of using longitudinal micro-level data for research, but it can also enormously expand the scope of future historical demographic research.

Once researchers decide to use IDS as a format, they are forced to rethink their data in terms of an international

standard by defining harmonized and unharmonized variables. This is a challenging process, regarding both time and cost. However, once complete, comparative studies become less time consuming and more cost-efficient. For the future of historical demography, this may be essential. In our work, we strive to understand why similar demographic transitions happened nearly at the same time in different societies, and to study how demographic behaviour across societies has changed over time. Today, individuals live close to each other, globally. Via internet, television marketing, and a growing number of global enterprises, we adopt a global behaviour in an increasingly integrated world. This change is mirrored in funding organizations, who are increasingly more willing to fund large comparative studies than studies based on one single country, e.g. Horizon 2020. The present demographic patterns need to be explained by cross-national comparisons of historical populations, and IDS makes this possible. The IDS structure brings together scholars by facilitating the establishment of research collaborations across different institutions as well as areas of knowledge. We hope that the current study is the first of many comparative works that will be conducted based on the IDS across different branches of historical demography.

REFERENCES

- 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(3), 78-114. doi: [10.12759/hsr.34.2009.3.78-114](https://doi.org/10.12759/hsr.34.2009.3.78-114)
- Bengtsson, T., Dribe, M., Quaranta, L., & Svensson, P. (2014). *The Scanian Economic Demographic Database. Version 4.0 (machine-readable database)*. Lund: Lund University, Centre for Economic Demography.
- Brändström, A. (1984). "De kärlekslösa mödrarna" spädbarnsdödligheten i Sverige under 1800-talet med särskilt hänsyn till Nedertorneå. Umeå: Almqvist & Wiksell.
- Broström, G., Edvinsson, S., & Engberg, E. (2018). Intergenerational transfers of infant mortality in 19th-century northern Sweden. *Historical Life Course Studies*, 7, 106-122. Retrieved from <http://hdl.handle.net/10622/23526343-2018-0005?locatt=view:master>
- Das Gupta, M. (1990). Death clustering, mother's 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)
- 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. Retrieved from <http://hdl.handle.net/10622/23526343-2018-0006?locatt=view:master>
- 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)
- 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)
- Mandemakers, K., & Laan, F. (2017). *LINKS dataset Genes Germs and Resources. WieWasWie Zeeland. Civil Certificates. Version 2017.01. IDS version. [Data file and code book]*
- 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. Available from <https://doi.org/10.1016/j.hisfam.2010.01.002>
- Puschmann, P. (2015). *Social inclusion and exclusion of urban in-migrants in northwestern European port cities. Antwerp, Rotterdam & Stockholm, ca. 1850-1930*. (Ph.D Thesis), University of Leuven
- 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. (2018a). Intergenerational transfers in infant mortality in southern Sweden, 1740-1968. *Historical Life Course Studies*, 7, 88-105. Retrieved from <http://hdl.handle.net/10622/23526343-2018-00013?locatt=view:master>

- Quaranta, L. (2018b). 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., Broström, G., Van Dijk, I., Donrovich, R., Edvinsson, S., Engberg, E., Mandemakers, K., Matthijs, K., Puschmann, P., & 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*.
- Reid, A. (2001). Neonatal mortality and stillbirths in early twentieth century Derbyshire, England. *Population Studies*, 55(3), 213-232. doi: [10.1080/00324720127696](https://doi.org/10.1080/00324720127696)
- Scalone, F., Agati, P., Angeli, A., & Donno, A. (2017). Exploring unobserved heterogeneity in perinatal and neonatal mortality risks: The case of an Italian sharecropping community, 1900–39. *Population Studies*, 71(1), 23-41. doi: [10.1080/00324728.2016.1254812](https://doi.org/10.1080/00324728.2016.1254812)
- Sommerseth, H. (2018). The intergenerational transfer of infant mortality in northern Norway during the 19th and early 20th centuries. *Historical Life Course Studies*, 7, 69-76. Retrieved from <http://hdl.handle.net/10622/23526343-2018-0008?locatt=view:master>
- Thorvaldsen, G., Andersen, T., & Sommerseth, H. L. (2015). Record linkage in the historical population register for Norway. In G. Bloothoof, P. Christen, K. Mandemakers & M. Schraagen (Eds.), *Population reconstruction*. Switzerland: Springer.
- Van den Boomen, N., & Ekamper, P. (2015). Denied their 'natural nourishment': Religion, causes of death and infant mortality in the Netherlands, 1875–1899. *The History of the Family*, 20(3), 391-419. doi: [10.1080/1081602X.2015.1022199](https://doi.org/10.1080/1081602X.2015.1022199)
- Van Dijk, I. K. (2018). Early-life mortality clustering in families: A literature review. *Population Studies*, 1-21. doi: [10.1080/00324728.2018.1448434](https://doi.org/10.1080/00324728.2018.1448434)
- 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. Retrieved from <http://hdl.handle.net/10622/23526343-2018-0003?locatt=view:master>
- Vandezande, M. (2012). *Born to die. Death clustering and the intergenerational transmission of infant mortality, the Antwerp district, 1846-1905*. (Ph.D thesis), University of Leuven.
- Westberg, A., Engberg, E. & Edvinsson, S. (2016). A Unique Source for Innovative Longitudinal Research: The POPLINK Database. *Historical Life Course Studies*, 3, 20-31. Retrieved from <http://hdl.handle.net/10622/23526343-2016-0003?locatt=view:master>