The Scanian Economic-Demographic Database (SEDD)

By Martin Dribe and Luciana Quaranta

To cite this article: Dribe, M. & Quaranta, L. (2020). The Scanian Economic-Demographic Database (SEDD). *Historical Life Course Studies*, Online first. https://hdl.handle.net/10622/23526343-2020-0008

HISTORICAL LIFE COURSE STUDIES

Content, Design and Structure of Major Databases with Historical Longitudinal Population Data

VOLUME 11 SPECIAL ISSUE 2, 2020

GUEST EDITORS
George Alter
Kees Mandemakers
Hélène Vézina

EHPS NETWORK

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) & 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.

HISTORICAL LIFE COURSE STUDIES
ONLINE FIRST (2020), published 12-11-2020

The Scanian Economic-Demographic Database (SEDD)

Martin Dribe
Lund University, Sweden

Luciana Quaranta
Lund University, Sweden

ABSTRACT

The Scanian Economic-Demographic Database (SEDD) is a high-quality longitudinal data resource spanning the period 1646–1967. It covers all individuals born in or migrated to the city of Landskrona and five rural parishes in western Scania in southern Sweden. The entire population present in the area is fully covered after 1813. At the individual level, SEDD combines various demographic and socioeconomic records, including causes of death, place of birth and geographic data on the place of residence within a parish. At the family level, the data contain a combination of demographic records and information on occupation, landholding and income. The data for 1813–1967 was structured in the model of the Intermediate Data Structure (IDS). In addition to storing source data in the SEDD IDS tables, a wide range of individual- and context-level variables were constructed, which means that most types of analyses using SEDD can be conducted without the need of further elaboration of the data. This article discusses the source material, linkage methods, and structure of the database.

Keywords: Sweden, Historical demography, Family reconstitution, Population registers, Longitudinal data, Life courses, Intermediate Data Structure

e-ISSN: 2352-6343
PID article: https://hdl.handle.net/10622/23526343-2020-0008
The article can be downloaded from here.

1 INTRODUCTION

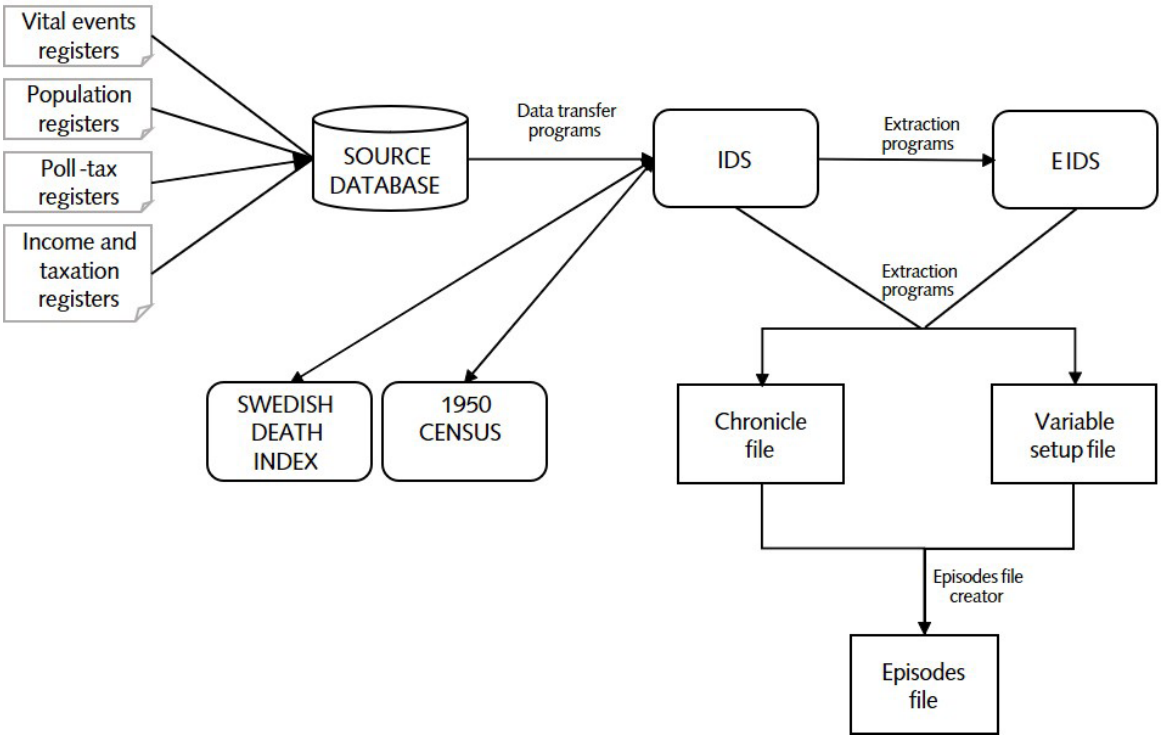
The Scanian Economic-Demographic Database (SEDD) is a high-quality longitudinal data resource spanning the period 1646–1967. It covers all individuals born in or migrated to the city of Landskrona and five rural parishes in western Scania in southern Sweden. Within a specific research program at the Center for Economic Demography at Lund University, individuals surviving to at least 1947 and their descendants have been linked to national registers from Statistics Sweden (*Statistiska centralbyrån*), the National Board of Health and Welfare (*Socialstyrelsen*) and the Military Archives (*Krigsarkivet, Pliktverket*), providing data also from 1968 up to present times.

SEDD is unique in several respects. It covers a longer period than most comparable databases and has a wealth of information at varying levels of aggregation. At the individual level, SEDD combines various demographic and socioeconomic records, including causes of death, place of birth and geographic data on the place of residence within a parish. The entire population present in the area is fully covered after 1813. For the city of Landskrona, the data cover the period 1905–1967, but the aim of an ongoing project is to complete the data entry back to 1880. At the family level, the data contain a combination of demographic records and information on occupation, landholding and income.

The multigenerational perspective of SEDD and the detailed information on a large number of variables are of great value to improve our knowledge of societal change in a broad sense. That SEDD covers the full period during which Sweden underwent enormous structural transformations, going from a preindustrial rural society to a modern welfare state, further underlining its scientific potential. SEDD is part of the Swedish national research infrastructure SwedPop, which aims to publish harmonized data from five different historical population databases using a common data structure (www.swedpop.se).

In this article, we first describe the source material and the linkage methods and then outline the data structure and variables in the database. The description covers the historical data up to 1968, and not the data from the national registers thereafter. Moreover, the data before 1813 will be covered in less detail as they are not included in the main database at present, and not structured in exactly the same way. These data are based on family reconstitutions, linked to poll-tax registers, rather than on population registers covering the entire resident population. A summary of the source material and the structure of the SEDD database is shown in Figure 1.

Figure 1 *Structure of SEDD*

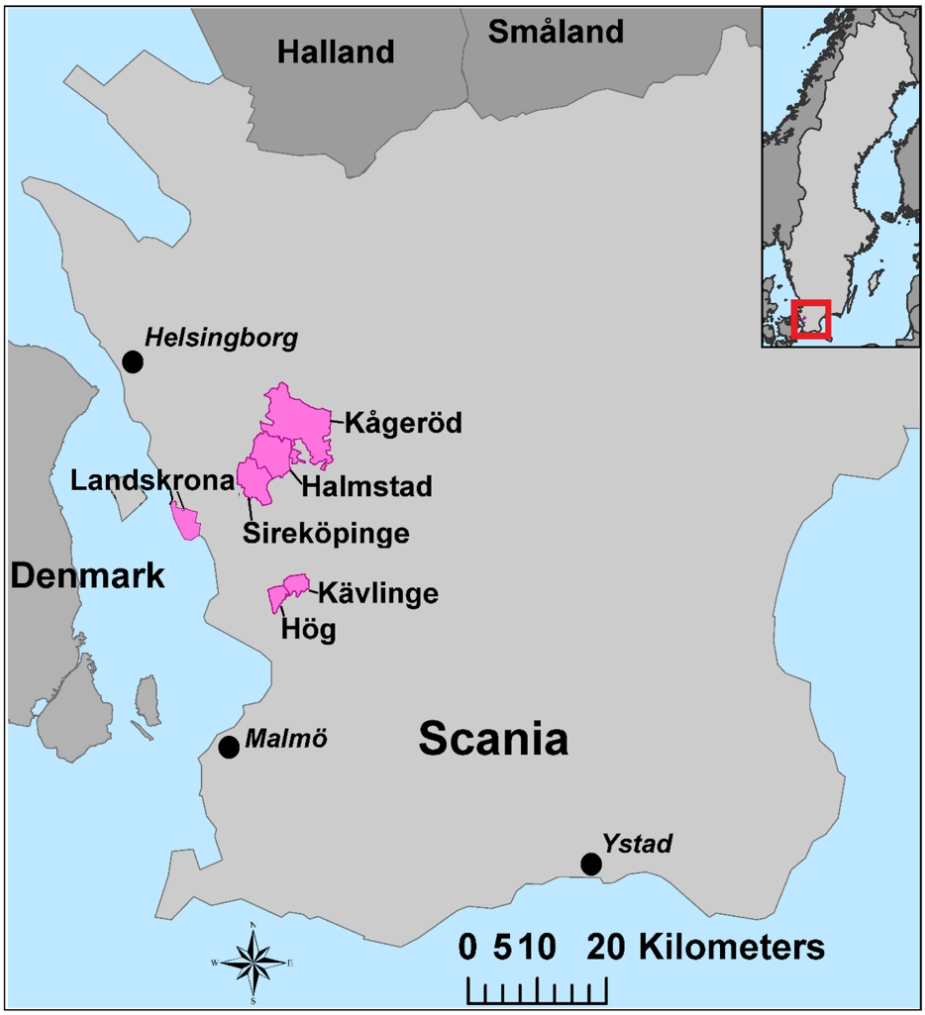


Source: Partly based on Figure 1 in Quaranta (2015).

2 THE AREA

SEDD consists of data from five rural and semi-urban parishes (Halmstad, Hög, Kågeröd, Kävlinge and Sireköpinge) and a port town, Landskrona, in Southern Sweden (Bengtsson, Dribe, Quaranta, & Svensson, 2020). The five parishes had a combined population of 4,300 in 1850, 5,900 in 1900 and, together with Landskrona, about 32,000 in 1950 (see figure 2 and table 1). For more detailed descriptions of the area and its socioeconomic and demographic development, see Bengtsson (2004), Dribe (2000), and Dribe and Svensson (2019).

Figure 2 *Map of SEDD areas*



Source: Map by Finn Hedefalk, Lund University.

Table 1 *Population totals in the parishes in SEDD*

	Hög	Kävlinge	Halmstad	Sireköpinge	Kågeröd	Landskrona
1750	208	193	312	360	1,274	
1805	292	344	529	613	1,412	
1850	550	631	787	817	1,544	4,139
1900	453	1,755	780	1,287	1,650	14,076
1950	365	3,208	348	881	1,796	25,089
1990	256	6,122	169	729	1,964	26,472

Sources: 1750–1850: Tabellverket, Umeå University (http://rystad.ddb.umu.se:8080/Tabellverket/Tabverk); 1900: BiSOS A, part 2, tab. 1; 1950: Statistics Sweden 1977 (Sveriges officiella statistik), Folkmängd 31 december 1950–1975; 1990: Statistics Sweden 1991 (Sveriges officiella statistik), Folkmängd 31 december 1990, del 1–2.

The parishes were chosen so that a geographically compact area could be obtained. This enable a study of the economic-demographic interactions without introducing biases stemming from, for example, regional differences. On the other hand, it will be difficult, without further qualifications, to generalize the results to cover Sweden as a whole.

Halmstad, Sireköpinge and Kågeröd are neighboring parishes, and were completely dominated by manorial (noble) land. They are located in a rather hilly part of the area, on the border between the agricultural plains and the more forested area of the northwest. Sireköpinge was plain land, Halmstad brushwood with some plain lands in the south and wooded areas in the north, and Kågeröd was more of a forest region. Hög and Kävlinge, on the other hand, are located on the plains, and here arable land dominated entirely, with virtually no wooded parts at all. In these two parishes landownership was dominated by freeholds and crown land. Thus, the topographical conditions as well as landownership differed considerably between the parishes. Towards the end of the 19th century, Kävlinge expanded considerably and was transformed into a small town as a result of the construction of the railroad and the establishment of a number of small industries.

Landskrona was founded in 1413 as a mercantile port town with a deep, natural harbor (see Dribe & Svensson, 2019). Later in the 16th century it also became an important fortified military town, but soon lost most of that role after Sweden gained control of the province of Scania. From the mid-19th century, factories and financial institutions were established. Its development was similar to other industrial cities, exemplified by the emergence of newspapers, schools, a hospital, institutional poor relief and old age care, a municipal board and governance, and a connection to the railway lines. The port was used for shipping grain from its hinterland, supporting the region's role as the country's breadbasket, and the last quarter of the 19th century saw the founding of mechanical factories and a shipyard, the latter of which would come to play an important role for the economy and also the identity of the town for nearly a century. In 1900, the population was about 14,000 and ranked as the tenth largest Swedish industrial city. The economic and demographic expansion continued until the industrial crisis of the 1970s, which had a large impact on the city.

3 CONTENT OF THE DATABASE

3.1 SOURCE MATERIAL

The core source material on which SEDD is based varies somewhat over time:

1680–1812 (only the five parishes):

- Family reconstitutions based on vital events (births, deaths, marriages)[1]
- Poll-tax registers (from 1766 annual coverage, before that year only partial)

1813–1967 (for Landskrona 1905–1967):

- Catechetical examination registers/parish books — population registers[2]
- Vital events registers (births, deaths, marriages)
- In- and out-migration registers
- Poll-tax registers (until 1945)

1862–1967 (for Landskrona 1905–1967):

- Income and taxation registers

In the following sections, we briefly discuss these different source materials. For parts of the database, information on heights, health and cognitive ability have been linked from muster rolls (Öberg, 2014)

1 In the earliest registers data on births come from baptisms and data on deaths come from burials. Data on reported births and deaths are available from 1685 in Hög and Kävlinge, 1710 in Halmstad and Sireköpinge, and from 1730 in Kågeröd.

2 The starting dates for the catechetical examination registers varies somewhat between the parishes: Kågeröd: 1813; Halmstad and Sirekpinge: 1821; and Hög and Kävlinge: 1829.

and detailed information of mother's and child's health from midwives' reports and contemporary birth registers (Lazuka, 2017). These sources will not be described in detail here but specific documentation is available for users of SEDD.

3.2 FAMILY RECONSTITUTIONS[3]

Family reconstitutions were carried out for the five parishes for the period 1646–1895. The method of family reconstitution has been widely used in historical demographic research since the 1950s, especially to reveal demographic patterns during periods for which no censuses or population registers are available. Furthermore, family reconstitutions are also a highly valuable complement to aggregate demographic data, when trying to reveal demographic patterns at the micro level (e.g. Lundh & Bengtsson, 1989). By linking information on births (sometimes baptisms), deaths (sometimes burials) and marriages, whole families can be reconstructed (Alter, 2020). Since the method is very laborious and time consuming, it is difficult to cover large areas. Hence, family reconstitution studies often cover only one or a few parishes.

Several problems and limitations of the method have been identified. Firstly, the church records suffer from a certain degree of under-recording of vital events. However, the Swedish marriage records seem to compare fairly well with those of some other European countries, which is mainly attributable to its religious homogeneity, which meant that most people married in the state church and were registered in the church records (Lundh & Bengtsson, 1989). However, regarding infant deaths there seems to be a problem of under-recording in the Swedish registers, so that some children that died shortly after birth were neither registered as born nor as dead (Bengtsson, 1999; Bengtsson & Lundh, 1994). Naturally, this would lead to underestimation of both marital fertility and infant mortality.

Secondly, there is a problem of tracing families that move between parishes. Such families will spread their demographic events in different parishes, which makes a correct estimation of their time of exposure difficult. Often, the solution to this problem has been to limit the analysis, of for example fertility, to the families that can be completely reconstructed, i.e. those that married in the parish and then stayed there for the whole of their reproductive period (Alter, 2020). Due to the rather high mobility in preindustrial society, this implies that only a minority of the families can be used; sometimes as few as 10–15% of the total population. Furthermore, there has been concerns that the families that did not move for such long periods were not representative of the population as a whole (e.g. Åkerman, 1977; Hollingsworth, 1969; Thestrup, 1972). In practice, however, the effect of this type of selectivity bias seems to be rather limited (e.g. Levine, 1976; Rogers, 1988). It has also been argued that even if 'stayers' really were representative of the population as a whole, estimates based on family reconstitutions could be biased, since the probability of moving before marriage was highest for those that married late. Hence, late marriages are likely to be under-recorded due to migration (Ruggles, 1992). Similarly, the probability of moving before death would be higher for those that live longer. Therefore, they will be under-represented in the population, which in turn would lead to underestimation of life expectancy. However, regarding age at marriage in England, Wrigley (1994) argued that the problem in practice was less serious than Ruggles claimed (see also Ruggles, 1999). In SEDD, where family reconstitutions are supplemented with data on migration, place of residence etc., this problem is much less of a concern.

Thirdly, there is also a problem of identifying the individuals and correctly linking them to the right family. In SEDD, a computer-based linking procedure, using standardized names, was used for the family reconstitutions (Bengtsson & Lundh, 1993). Overall, the performance of the program was quite satisfactory: 96% of the births, 72% of the marriages and 55% of the deaths were correctly linked by the program (Bengtsson & Lundh, 1991). Since then, manual corrections raised these figures considerably (99% of births, 100% of marriages and 90% of the deaths were linked to families for the period before 1800).

3.3 POLL-TAX REGISTERS

Poll-tax registers (*mantalslängder*) are available from the late 17th century until 1945 and have been used to obtain information on where the families lived, and whether they had access to land or not. They were yearly registers, used for collecting taxes and containing information on the size of the

3 The sections on family reconstitutions, poll-tax registers and population registers are based on Dribe (2000, chapter 2).

landholding, the type of ownership (i.e. noble, crown, church or freehold) and information on the number of servants and lodgers. In addition to the poll-tax registers, land registers (*jordeböcker*) have also been utilized to clarify the ownership of land, when poll-tax registers were missing. Information from these two registers was linked to the reconstituted families. Thereby, information was obtained not only on the demographic events, but also on the economic realities of these families. After 1766 the registers are available on an annual basis and linkage between the poll-tax registers and the family reconstitution is almost entirely complete (i.e. only odd holdings that have not been linked to a family).

In the poll-tax registers, the size of the landholding is expressed in *mantal*. This was an old tax unit, originally meaning 'the number of men'. At the beginning, during the 16th century, every landholding was supposed to have one *mantal*, i.e. be large enough to support one peasant and his family as well as producing a surplus to be paid as tax to the crown (Hechscher, 1949). Thus, at this time a *mantal* only meant that the peasant had land and was supposed to pay tax to the crown. However, due to repeated subdivisions of landholdings, farmsteads typically got smaller and smaller fractions of a *mantal* assigned to them. Furthermore, reclamation of new land, as well as changed methods of cultivation lead to increased land productivity, which makes a comparison over time of the size of different farms almost impossible. Nevertheless, the *mantal* can be used at least as a rough measure of the size of a farm relative to other farms in the village at the same point in time. Thus, by comparing the different *mantal* peasants had, the relative sizes of the landholdings could be determined.

The *mantal* assigned to a farm was based on the quantity of seed sown and hay harvested, which in turn reflected the productive potential of the arable land and of the meadow, even though other concerns were taken as well, such as the productive strength of underlying crofts or cottages (Sommarin, 1939). In western Scania in 1820, one *mantal* corresponded approximately to 180–230 acres (150–190 *tunnland*). At the same time, the minimum amount of land required to be a self-sufficient farmer (*besuttenhetsgräns*) was 1/16 of a *mantal*, which corresponded roughly to 15 acres (Sommarin, 1939).

Besides assessing the economic status of families, the poll-tax registers play a crucial role in determining the time of exposure for the families. By linking the reconstituted families to the place where they lived, we get the time they spent in the parish, and thereby we also get a rather precise estimate of time that these families are under observation. This information is of great use particularly for pre-1813 data, before population registers were available. It is important to note, however, that the information refers to the family, and not to individuals.

3.4 POPULATION REGISTERS (CATECHETICAL EXAMINATION REGISTERS AND PARISH BOOKS)

During the 18th century, a regulation concerning examinations of the biblical knowledge of the parishioners was introduced. To begin with, the examinations were held in church, by the people living in the parish coming to church (*kyrkoförhör*). Later, the examinations were instead conducted in people's homes (*husförhör*). The clergymen needed a list of the inhabitants in order to conduct the examinations, which was the origin of the catechetical examination registers. With the establishment of the Tabular Commission (*Tabellverket*) in 1749, the catechetical examination registers became the basis for the parish level records (Lext, 1984). This made it compulsory to record information on births, deaths and migration in the registers. In addition, the registers sometimes included information on reading and writing ability, smallpox vaccination, and various comments made by the clergy regarding the inhabitants. In the poll-tax instruction of 1812, the catechetical examination registers also got a very important role in controlling the poll-tax registers (Lext, 1967, 1984).

Every head of household was obliged to report the members of his household and the people living at his holding at the time of examination. Over the 19th century, the connection between the actual catechetical examinations and the catechetical examination registers diminished, and the clerical role of the registers disappeared to a large extent. Instead the demographic and administrative role of the registers became increasingly important (Lext, 1984), but the registers were still the responsibility of the parish. From 1895, the registers were renamed parish books (*församlingsböcker*) but contained the same basic information as before. This system of population registration was in place until 1991, when it was transferred to the tax authorities and became centralized.

Overall, there is a high degree of correspondence between the events in the catechetical examination registers and the records of vital events. However, children that died in infancy sometimes appear to

have been poorly registered in the catechetical examination registers, because not all children were entered immediately upon birth (see also Winberg, 1975).

In the population registers, the families and households were recorded according to where they lived. People belonging to the same household were listed together. Usually, it is relatively easy to identify to which household the head of household's family and his servants belonged based on where they are entered in the registers. However, it is sometimes more difficult to identify the right household for lodgers and retired people.[4] If they are registered together with a certain family at a farm they have been included in that household, but in those cases where it has been difficult to determine to which household they belonged, they have been registered in a household of their own. As has been pointed out by other researchers, it is impossible to use the catechetical examination registers as a basis for a theoretical definition of a household (Gaunt, 1977). The definition used in SEDD focuses on the household as a production and consumption unit, and includes servants, lodgers and retired people in addition to the nuclear family of the head of household. This means that all the people living at the farm, contributing to production and/or dependent on the household for consumption, are included in the household, while temporary salaried workers that did not live in the household are not included. From 1947 and onwards, the population registers include unique personal identifiers given to all Swedish residents (*personnummer*).

3.5 INCOME AND TAXATION REGISTERS

For the period 1862–1967, SEDD contains information about income and taxes paid, derived from the income and taxation registers (*inkomst- och taxeringslängderna*). For the five parishes the information is available annually for the whole period 1862–1867, while for Landskrona it is available annually from 1947–1967, and about every five years from 1905–1946. Work is currently ongoing to complete the registration of income registers for all years between 1904 and 1946. While incomes initially were assessed by the local tax authorities, from 1902 onwards it was based on tax returns made by the tax payers themselves (*självdeklarationer*). The exact information included in the registers varies substantially over time, and changing income thresholds in the tax law implies that an increasing share of the population is included in the registers. Helgertz, Dribe, and Bengtsson (2020) provide a detailed description of the source material and the income variables included in SEDD.

Until 1971, Sweden practiced joint taxation of married couples, but after 1947 income of married women were nonetheless reported separately in the registers. Before 1947, married women are not included in the tax registers, but their incomes are reported together with that of their husbands.

3.6 LINKS TO THE SWEDISH DEATH INDEX AND THE CENSUS OF 1950

SEDD has been linked to both the Swedish Death Index (SDI) and the Census of 1950. The SDI has been developed by the Federation of Swedish Genealogical Societies (2019) and includes most deaths in Sweden between 1860 and 2017. After 1947, the index also includes the unique personal identification numbers which allow for a direct link to SEDD. Before 1947, the sources were linked using automated matching based on names and date and place of birth. The links are valuable in providing information about the time and place of death for out-migrants, as well as personal identification numbers for people moving out before 1947 (when we have the personal identifiers in SEDD), but dying after 1968 so that they could be traced in the national registers before they died.

In a similar way, the Census of 1950 (digitized and transcribed by Arkiv Digital, www.arkivdigital.se) was linked to SEDD using either personal identification numbers or through automatic matching based on names and date and place of birth. The census provides snapshot information from 1950 on occupation, family context and place of residence and, more importantly, gives personal identification numbers for people leaving the area before 1947, allowing more SEDD individuals to be linked to the national registers.

4 Since leasing agreements are not fully captured in the poll-tax registers, it is difficult to know which families lived at the same farm as each other and thus shared environment. For example, when several smaller units are listed under the main farm it can be difficult (but possible using historical maps) to link the correct tenant to the unit they rent, but even more difficult to link other families in the poll-tax registers that resided in either the rented unit or in the 'main' property unit, but in different households than the tenant/owner.

3.7 FATHER'S OCCUPATION AT BIRTH OF THE CHILD

All adults present in SEDD have been traced back to their parish of birth to retrieve information on the occupation of the father. The information has been taken from the birth register, or in some cases the parish book. This means that the database contains information on father's occupation not only for children born in the parishes, but for all individuals who were present in SEDD as adults and for whom that information could be found in the parish of birth.

4 DATA CODING AND LINKAGE BETWEEN SOURCES

4.1 CODING OF OCCUPATIONS

All occupations have been coded into HISCO, the *Historical International Standard Classification of Occupations* (van Leeuwen, Maas, & Miles, 2002). HISCO is a historical classification based on the International Labour Organisation's classification ISCO68 (ILO, 1969) and attempts to increase the comparability of historical occupational titles. In HISCO occupations are categorized according to tasks that need to be fulfilled in that occupation. HISCO divides occupations into eight major groups (e.g. major group 5 'Service Workers'), each of which is divided in two to ten minor groups (e.g. minor group 5.3. 'Cooks, Waiters, Bartenders and Related Workers'). These 83 minor groups are again subdivided into 284 unit groups (e.g. 5.31 'Cooks'). Finally, these unit groups consist of 1,881 occupational categories, the lowest level of detail (e.g. 5.31.50 'Ship's Cook') (see Dribe, Helgertz, & Van de Putte, 2015). Occupations with comparable tasks are grouped into one of these categories. Initially the coding was done manually, but in the present version of SEDD the occupational coding done in the SwedPop project has been applied to the SEDD data. This coding is a harmonized version of HISCO coding done at different historical population databases in Sweden. Based on HISCO codes, social class schemes, such as HISCLASS (Historical International Social Class Scheme, see van Leeuwen & Maas, 2011), and rank schemes, such as HISCAM (Lambert, Zijdeman, van Leeuwen, Maas, & Prandy, 2014), can be easily applied using existing transcode tables.

4.2 CODING OF CAUSES OF DEATH

SEDD includes information about cause of death. The unique text strings indicating the causes listed in the death registers have been coded into ICD10 (*International Statistical Classification of Diseases and Related Health Problems*) within the SwedPop project. As with the HISCO codes, this is a harmonized coding of causes of death applied in all the large historical population databases in Sweden (Hiltunen & Edvinsson, 2018).

4.3 DATA LINKAGE 1813–1967

The data from the different sources have been linked together manually, in most cases when a new source was added to the database. Information about name, date of birth, place of birth, and family context have been important in identification and linking of records. The vital events registers have also been linked to the population registers to check for under-recording of events, and to add missing data when necessary.

As already mentioned, linking of SEDD to the census of 1950 and the SDI, as well as the original family reconstitution of the five parishes, was done wholly or partly using automatic linkage methods.

Table 2 shows the number of events and unique individuals in the database 1813–1967. In total, it includes about 175,000 unique individuals, of which 74,000 have personal identification numbers that allow then to be linked to the national contemporary registers. The database contain about 43,000 births, 27,000 deaths and almost 300,000 migrations, out of which about 20,000 moves are between the parishes in the database.

Table 2 *Number of events, individuals, families and households in SEDD, 1813–1967*

	N
Births	43,129
Deaths	26,676
Marriages	26,171
In-migrations	148,831
Out-migrations	145,755
Number of individuals	175,149
Individuals with national id number	74,213
Number of family IDs	116,888
Number of household IDs	79,744

Note: In/out migrations are moves over parish borders, also including migrations between two parishes in the sample.

4.4 GEOCODING[5]

Through geocoding, the place of residence is known for parts of the population in the five parishes and Landskrona at various geographic levels and periods.

For the five parishes, the approximately 53,000 residents[6] for the period 1813–1914 have been linked to the property units at which they lived (see Hedefalk, Harrie, & Svensson, 2015). To enable such geocoding, an object-lifeline representation of the digitized property units was created (using historical maps and poll-tax registers), which contains information about when the units were created, changed, and ceased to exist. Such a representation was necessary to accurately trace the residential histories in time, as the property unit borders were often subdivided or partitioned into smaller units in line with the rapid population growth during the period. Thus, each move within the parishes can be traced. In addition to the geocoding, historical wetlands (in object-lifelines), roads, buildings and water in the five parishes were digitized for the same period (see Hedefalk et al. (2017) for details on the geocoding match rate and datasets).

For Landskrona, the full population has been geocoded at: (1) block-level for the period 1904–1938; and (2) address-level for the period 1939–1967. Thus, continuous information about the individual place of residence at a very detailed geographic level is available. For the period 1939–1967, individuals are linked to the exact addresses they lived at, and each move is traced. For the period 1904–1938, the geocoding has been performed on the slightly coarser block-level (Swedish: *kvarter*). In addition, an object-lifeline representation of the buildings and streets in Landskrona was created for the period 1904–1967, allowing to link individuals to the buildings they resided in (more accurately for the period 1939–1967; see Hedefalk and Dribe (2020) for information on geocoding match rate).

5 We thank Finn Hedefalk for contributing to this section.

6 For the geocoding of the five parishes, we primarily used information from the poll-tax registers, which contains the addresses (property units) of each person who had to pay taxes. The poll-tax registers are on a household level, which indicates that only the head of the household is noted with a full name. However, each individual in the SEDD is linked to the families and households to which they belong, and most of these families and households have a correspondence in the poll-tax registers.

5 DATA STRUCTURE AND VARIABLES

5.1 FROM SOURCE MATERIAL TO A DATASET FOR ANALYSIS

The linked data from the different source registers is stored in a relational database in Microsoft SQL server: the SEDD source database (see Figure 1). This database is composed of different tables containing information from each source, and tables with standardized variables (occupations, locations and causes of death). The SEDD source database is continuously updated as new data is digitized or any information is changed.

The data for 1813–1967 was structured in the model of the Intermediate Data Structure (IDS).[7] The IDS was introduced with the aim of creating a common format for the standardization and dissemination of data from different databases, regardless of their original form (Alter & Mandemakers, 2014). It also provides standardized solutions for storing constructed variables, making data extractions and preparing datasets for analysis, and allows to develop and share common software (Quaranta, 2015). The IDS is not meant to replace source databases, but rather to serve as a middle layer between the source database and rectangular datasets for analysis.

IDS databases are composed of two types of entities, persons and contexts, and the relation among persons, contexts and between persons and contexts (Alter & Mandemakers, 2014). Contexts represent physical or social spaces that group individuals together. Data is stored in the IDS tables following the entity-attribute-value model, which contains one 'attribute' (i.e. variable) per record. Each record includes a *Type* or *Relation,* indicating the kind of information it contains, which can be the date of an event or *Value* of an attribute, or the type of relationship that exists between two persons or between a person and a context (Alter, Newton, & Oeppen, 2020). Each record also includes an identifier of the person (*Id_I*) or context (*Id_C*). Attributes are dated using a *Time Stamp*.

To conduct longitudinal statistical analysis using data stored in IDS, data extractions must be transformed into rectangular data arrays, also called 'episodes tables'. Episodes are spells of time during which the values of attributes remain constant and at the end of which the event of interest of the study can take place. The start and end dates of the rows of an episodes table correspond to the dates when any of the attributes or events included in the data extraction change value. The transformation of IDS data into episodes tables can be done using the two-step procedure proposed in Quaranta (2015). In the first step, a Chronicle and a Variable setup file are produced. The Chronicle file contains all individual and contextual level attributes and all events selected by the researcher for analysis. Each *Type* of attribute included in the Chronicle file becomes a column of the episodes table. In the second step, episodes tables can be automatically made from these two files, using the STATA program 'Episodes file creator' (Quaranta, 2016). The Variable Setup file stores information relating to each attribute included in the Chronicle file, and this information is used by the 'Episodes file creator' to identify how each attribute should be treated when creating the episodes table.

5.2 THE SEDD IDS TABLES

The IDS consists of five main tables: INDIVIDUAL storing information relating to individuals; INDIV_INDIV defining relationships between individuals; CONTEXT defining contexts and storing information about them; CONTEXT_CONTEXT defining contexts that are nested in other higher level contexts; and INDIV_CONTEXT identifying spells of times during which individuals are present in a specific context (Alter & Mandemakers, 2014). A METADATA table is also used, which defines all attributes included in the tables and their values.

The SEDD INDIVIDUAL table includes many attributes. Dates of birth and marriage are available, and for such records the *Date_type* field of the table specifies whether the actual birth and marriage events were observed or whether their dates were declared in other sources. Birth and marriage locations and information on death date, location and cause of death (ICD10 codes) are also stored in the table.

7 Separate IDS tables were made for the family reconstitution data, which include source information from the vital event registers and the poll-tax registers. Family reconstitution rules should be adopted when using this data for research.

Additional variables include sex, arrival, departure and start and end observations.[8] The INDIVIDUAL table also stores occupations of individuals, indicating in the *Source* field of the table which register the occupation comes from (parish, or income and taxation registers), and the occupation of the father at the time of birth of the individual. In both cases, standardized strings and HISCO coding are available. The types of relations included in the SEDD INDIV_INDIV table are husband, wife, child, father, mother, adoptive child/father/mother, foster child/father/mother and step child/father/mother.[9]

In the SEDD, four different levels are included in the CONTEXT table: family, household, parish and poll-tax.[10] The family includes all related persons within one household. It corresponds to the nuclear family, also referred to as conjugal family unit (CFU — see e.g. Hammel & Laslett, 1974). CFUs may be formed in any of the following ways: by a couple without offspring; by a couple with unmarried offspring and/or unmarried adopted/foster children; by a lone parent with at least one never-married child; or by a single adult. Families never include more than two generations. If more than two generations live in the same household, the CFU is formed from the youngest generation upwards.[11] Servants and lodgers constitute their own families. Individuals are linked to family contexts in the INDIV_CONTEXT table, specifying the type of relation that they have to the family (man, woman, child). Individuals can only belong to one family at any specific point in time, but they can move across families (e.g. family of birth and family of marriage).

As discussed above, households in the SEDD include the nuclear family of the head of household as well as servants, lodgers and retired people. Several families may live in the same household. For example, a household will host the main family (which is the head of household, his wife and children), plus a family of servants, consisting of a married couple and their children. Individuals are linked to a household context in the INDIV_CONTEXT table, specifying the type of relation that they have to the household (householder, relative, boarder, servant or unknown). Individuals can only belong to one household at any specific point in time, but they can move across households. In SEDD families are not nested within households since it is possible for individuals who belong to the same family to live in different households at the same point in time.[12] Households are linked to parishes through the CONTEXT_CONTEXT table.

A specific context layer was created for poll-tax, to allow to more easily link information from these registers to all members of the family. The variables from poll-tax registers that are stored in the CONTEXT table are holding and land type, size fraction (*mantal*) and occupation. Poll-tax registers are linked to families through the CONTEX_CONTEXT table. Since families could own more than one property, it is possible for them to have more than one link to a poll-tax register during the same year.

5.3 THE SEDD EXTENDED IDS TABLES

In addition to storing source data in the SEDD IDS tables, a wide range of individual- and context-level variables were constructed using source information, like household size, total number of children in the family, occupation of the family head, etc. They were stored in the extended IDS tables (EIDS — Quaranta, 2015) INDIVIDUAL_EXT and CONTEXT_EXT. Most types of analyses using SEDD can therefore be conducted without the need of further elaboration of the data. This not only facilitates the work of researchers, but it also makes different research outputs more consistent, since variables are defined identically.

8 All dates of migration to and from SEDD areas and within SEDD areas are recorded. When it was possible, individuals who moved between different SEDD parishes were linked and given unique identifiers. Such links could be made for all individuals after 1947, when personal identification numbers became available. In preceding years, links were established if there was a full match on name and date and place of birth; however, this procedure is unlikely to have captured all cases.

9 The relations stored in the INDIV_INDIV table correspond to those found in the source material. Other types of kinship relations can be identified using the information from this table. For example, if individual 1 is the mother of individual 2, the grandchildren of 1 can be identified by selecting from the table all children linked to 2.

10 In the family reconstitution IDS dataset, the CONTEXT table includes the levels family and poll-tax, but not household, given that population registers are not available.

11 Other extended family members, such as siblings of the head of family also constitute their own separate family.

12 An example can be a married woman spending a period as lodger in a different household and later returning to the household where the rest of her family lives.

The CONTEXT_EXT table stores constructed context-level variables. Two of such variables are family head ID and household size. The occupations of the family head are also included in the tables, which correspond to the occupations, from the parish registers and from the income and taxation registers, of the individual defined as the family head. The source field of variables is not retained when constructing an episodes table. However, because of the importance in research to distinguish occupations from different sources, all created extended occupational variables incorporate the source into the variable name. Amongst such created variables are the occupations from the poll-tax registers, and since families can have more than one poll-tax declaration each year, up to three poll-tax occupations may be available. Three variables indicating the date when occupations were last declared are also stored in the table; one for the occupations of the family head from the parish registers, one for the occupations of the family head from the income and taxation registers and one for the occupations from the poll-tax registers. When constructing the episodes tables, values of occupations are copied down until the next time an occupation is declared. Using the declaration date variables, researchers can later decide for how long to consider occupations valid.[13]

The INDIVIDUAL_EXT table contains currently contains more than 100 variables, which can be divided into different groups. One group of variables measures characteristics of the family, some of which only relate to children (mother presence, father presence, older sisters, older brothers, younger sisters, younger brothers), others only to the man and the woman (number of children), and others to all members of the family (proportion of previously born children who are dead, number of surviving daughters, number of surviving sons). There is also one general individual level variable, civil status, which takes into consideration whether the individual has a partner, and whether the partner is present in the same household.

The variables stored in the CONTEXT_EXT table (household size, size fraction, holding type, land type, poll-tax occupations and occupations of the family head) could also be assigned to individuals for the specific periods of time in which they were present in the relative context. Such information is stored in the INDIVIDUAL_EXT table as well. This makes this information redundant in the database, but it has the advantage that it can more easily be used in research. Occupations of the individual, from the parish registers and from the income and taxation registers, are also stored in INDIVIDUAL_EXT, incorporating also in this case the source information into the name of the variable. Since each occupational title may contain up to four occupations, for each occupational variables there are twelve variables related to HISCO coding (code, relation and status). Moreover, the table stores variables with the dates of declaration of each type of occupational variable (see above).

Finally, the INDIVIDUAL_EXT table also stores a series of variables that are specific to mortality or fertility research, and a variable measuring whether the individual was present in the study area, which is essential for most kinds of longitudinal research. For mortality research, it includes an indicator variable of whether the individual died at the end of the spell. For fertility research, the table includes several variables, which are assigned only to females: indicator of child birth, date of previous child birth, date of previous marriage, number of previous births, status of the previously born child.

13 For example, the occupation of the family head, from the parish registers, could have the value farmer, declared on March 13, 1856, when the family first enters the study area, the value shoemaker, declared on August 5, 1862, and no other occupational declarations until the family leaves the study area on January 21, 1865. When creating the episodes table the value farmer is assigned to the spell between March 13, 1856 and August 5, 1862, and the value shoemaker is assigned to the spell between August 5, 1862 and January 21, 1865. Since there are many variables in the database, which change values in different dates, each spell between declarations of occupations is generally split into more rows, making it difficult for researchers to know when an occupation was actually declared. This is why the variable indicating the date of declaration of occupations becomes important. Using the date variable a researcher could for example decide to hold occupations valid for up to five years and therefore replace the occupation value to missing between March 13, 1861 and August 5, 1862.

6 USING SEDD IN RESEARCH

In addition to the IDS and EIDS tables, a Chronicle file and a Variable setup file are provided to researchers.[14] The SEDD Chronicle file contains all source and constructed variables and events. As mentioned earlier, the Chronicle and Variable setup files can be automatically transformed into a rectangular episodes table that is ready for statistical analysis using a STATA program (Quaranta, 2016). The episodes table can also be exported to conduct analysis using other software. In addition to the variables provided, researchers can create other variables using information contained in the IDS tables or external data. Such variables should be included in the Chronicle and Variable setup files prior to creating the episodes table.

The transformation of the SEDD source database into IDS and the construction of variables for analysis is made using SQL queries in Microsoft Access and Microsoft SQL Server Management Studio. A new IDS database and new Chronicle and Variable setup files are produced every time there is a new SEDD release. New releases are made when sections of digitization are completed, or when substantial errors are fixed in the data.

Data from SEDD older than 100 years are publicly accessible without restrictions.[15] More recent data pertaining to individuals who are still alive can be released for scientific research subject to legal restrictions on the use of personal data (e.g. GDPR and the Swedish law of ethical review of research).

7 SUMMARY AND FUTURE PERSPECTIVES OF SEDD

This article has described the SEDD database, its sources and structure. As was shown, SEDD contains detailed and rich longitudinal demographic and socioeconomic data, which expand over a very long time period and across several generations. The conversion of SEDD into IDS has meant that the structure of the database is clear and well documented, making SEDD easily available for researchers. Many variables have also been constructed and episodes tables can be automatically created, meaning that most analyses can be made without additional elaboration of the data.

Even though SEDD is very rich, there are several plans to continue its expansion. As mentioned earlier, the data for Landskrona will be extended back to 1880. More socioeconomic and health variables will be added to SEDD, expanding further the scope of the data. For example, hospital birth records are available for children born between 1935 and 1945 and more cohorts are currently being digitized. Like many other longitudinal databases, the main limitation of SEDD is its geographical coverage. The IDS has however allowed to conduct fully comparative studies across different European populations (e.g. Quaranta, 2018; Quaranta & Sommerseth, 2018), increasing the generalizability of results. SEDD is also part of the SwedPop project, which aims to harmonize and disseminate historical population data from different sources and areas of Sweden (www.swedpop.se). Swedpop will vastly expand the possibilities to conduct comparative research considering other areas in Sweden.

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 Tommy Bengtsson and Finn Hedefalk for comments and suggestions.

14 In the IDS family reconstitution database, only basic IDS tables were made, i.e. no extended variables were constructed and the Chronicle and Variable Setup file are not provided.

15 For more information, see https://www.ed.lu.se/databases/sedd/sedd-public-access.

REFERENCES

Åkerman, S. (1977). An evaluation of the family reconstitution technique. *Scandinavian Economic History Review, 25*, 160–170. doi: 10.1080/03585522.1977.10407879

Alter, G. (2020). The evolution of models in historical demography. *Journal of Interdisciplinary History, 50*(3), 325–362. doi: 10.1162/jinh_a_01445

Alter, G., & Mandemakers, K. (2014). The Intermediate Data Structure (IDS) for longitudinal microdata, version 4. *Historical Life Course Studies, 1*(1), 1–29. Retrieved from http://hdl.handle.net/10622/23526343-2014-0001?locatt=view:master

Alter, G., Newton, G., & Oeppen, J. (2020). Re-introducing the Cambridge Group Family Reconstitutions. *Historical Life Course Studies, 9*, 24–48. Retrieved from https://hdl.handle.net/10622/23526343-2020-0005

Bengtsson, T. (1999). The vulnerable child. Economic insecurity and child mortality in preindustrial Sweden: A case study of Västanfors, 1757–1850. *European Journal of Population, 15*, 117–151. doi: 10.1023/A:1006215701608

Bengtsson, T. (2004). Mortality and social class in four Scanian parishes. In: Bengtsson, T., Campbell, C., & Lee, J. Z. (Eds.). *Life under pressure: Mortality and living standards in Europe and Asia 1700–1900* (pp. 37–41). Cambridge, MA: MIT Press.

Bengtsson, T., Dribe, M., Quaranta, L., & Svensson, P. (2020). *The Scanian Economic Demographic Database. Version 7.1 [machine-readable database]*. Lund: Lund University, Centre for Economic Demography.

Bengtsson, T., & Lundh, C. (1991). *Evaluation of a Swedish computer program for automatic family reconstitution*. (Lund Papers in Economic History, No. 8). Lund: Lund University, Department of Economic History.

Bengtsson, T., & Lundh, C. (1993). *Name-standardisation and automatic family reconstitution*. (Lund Papers in Economic History, No. 29). Lund: Lund University, Department of Economic History.

Bengtsson, T., & Lundh, C. (1994). Child and infant mortality in the Nordic countries. *Annales de démographie historique*, 24–43. doi: 10.3406/ADH.1994.1857

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., 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*, 25–39. doi: 10.1016/j.rssm.2015.04.005

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, Department of Economic History. Retrieved from https://www.ed.lu.se/media/ed/papers/working_papers/LPED_2019_3.pdf

Gaunt, D. (1977). Pre-industrial economy and population structure: The elements of variance in early modern Sweden. *Scandinavian Journal of History, 2*(1–4), 83–210. doi: 10.1080/03468757708578918

Hammel, E. A., & Laslett, P. (1974). Comparing household structure over time and between cultures. *Comparative studies in Society and History, 16*(1), 73–109. Available from https://doi.org/10.1017/S0010417500007362

Heckscher, E. F. (1949). *Sveriges ekonomiska historia från Gustav Vasa, del 2*. Stockholm: Albert Bonniers förlag.

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

Hedefalk, F., Harrie, L., & Svensson, P. (2015). Methods to create a longitudinal integrated demographic and geographic database on the micro-level: A case study of five Swedish rural parishes, 1813–1914. *Historical Methods: A Journal of Quantitative and Interdisciplinary History, 48*(3), 153–173. doi: 10.1080/01615440.2015.1016645

Hedefalk, F., Svensson, P., & Harrie, L. (2017). Spatiotemporal historical datasets at micro-level for geocoded individuals in five Swedish parishes, 1813–1914. *Scientific data, 4*(1), 1–13. doi: 10.1038/sdata.2017.46

Helgertz, J., Dribe, M., & Bengtsson, T. (2020). *Income data in the Scanian Economic Demographic Database (SEDD)*. (Lund Papers in Economic Demography, No. 6). Lund: Lund University, Department of Economic History. Retrieved from https://www.ed.lu.se/media/ed/papers/working_papers/LPED_2020_6.pdf

Hiltunen, M. & Edvinsson, S. (2018). Classifying literate cause-of-death information originating from Swedish historical parish registers. Unpublished paper.
Hollingsworth, T. H. (1969). *Historical Demography*. London: Hodder & Stoughton.
International Labor Office (ILO). 1969. *International standard classification of occupations, revised edition 1968*. Geneva: International Labor Office.
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
Lazuka, V. (2017). *Defeating disease. Lasting effects of public health and medical breakthroughs between 1880 and 1945 on health and income in Sweden* (PhD Thesis Economic History, Lund University). Retrieved from https://lup.lub.lu.se/search/publication/1cba7526-042f-43fe-9005-6f6bc8cbf862
Levine, D. (1976). The reliability of parochial registration and the representativeness of family reconstitution. *Population Studies, 30*(1), 107–120. doi: 10.1080/00324728.1976.10412723
Lext, G. (1967). *Mantalsskrivningen i Sverige före 1860*. (Meddelanden från Ekonomisk-historiska institutionen, No. 13). Göteborg: Ekonomisk-historiska institutionen.
Lext, G. (1984). *Studier i svensk kyrkobokföring 1600–1946*. (Meddelanden från Ekonomisk-historiska institutionen vid Göteborgs universitet, No. 54). Göteborg: Ekonomisk-historiska institutionen.
Lundh, C., & Bengtsson, T. (1989). *Familjerekonstruktion på svenskt kyrkoboksmaterial. Problem och möjligheter*. (Meddelande från ekonomisk-historiska institutionen, Lunds universitet, No. 59). Lund: Ekonomisk-historiska Institutionen.
Öberg, S. (2014). *Social bodies: Family and community level influences on height and weight, southern Sweden 1818–1968* (PhD Thesis Economic History, University of Gothenburg). Retrieved from https://www.researchgate.net/publication/281290743_Social_bodies_family_and_community_level_influences_on_height_and_weight_southern_Sweden_1818-1968
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). Intergenerational transfers in infant mortality in southern Sweden, 1740–1968. *Historical Life Course Studies, 7*, 88–105. Retrieved from http://hdl.handle.net/10622/23526343-2018-0013?locatt=view:master
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. Retrieved from http://hdl.handle.net/10622/23526343-2018-0014?locatt=view:master
Rogers, J. (1988). *Family reconstitution: new information or misinformation*. (Reports from the Family History Group, 7). Uppsala: Department of History, Uppsala University.
Ruggles, S. (1992). Migration, marriage and mortality: Correcting sources of bias in English family reconstitutions. *Population Studies, 46*(3), 507–522. doi: 10.1080/0032472031000146486
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
Sommarin, E. (1939). *Det skånska jordbrukets ekonomiska utveckling 1801–1914, del 2–3*. Lund: Skrifter utgivna av de skånska hushållningssällskapen.
The Federation of Swedish Genealogical Societies (2019). *Sveriges dödbok, version 7, 1860–2017* [Machine readable database].
Thestrup, P. (1972). Methodological problems of a family reconstitution study in a Danish rural parish before 1800. *Scandinavian Economic History Review, 20*(1), 1–26. doi: 10.1080/03585522.1972.10407708
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.
Winberg, C. (1975). *Folkökning och proletarisering. Kring den sociala strukturomvandlingen på Sveriges landsbygd under den agrara revolutionen* (PhD Thesis History, University of Gothenburg).
Wrigley, E. A. (1994). The effect of migration on the estimation of marriage age in family reconstitution studies. *Population Studies, 48*(1), 81–97. doi: 10.1080/0032472031000147486