

From Matched Certificates to Related Persons

By Rick Mourits, Ingrid K. van Dijk, & Kees Mandemakers

To cite this article: Mourits, R., van Dijk, I.K., & Mandemakers, K. (2020). From Matched Certificates to Related Persons. *Historical Life Course Studies*, 9, 49–68. <https://hdl.handle.net/10622/23526343-2020-0006>

HISTORICAL LIFE COURSE STUDIES

VOLUME 9

2020



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>.



From Matched Certificates to Related Persons

Rick Mourits

Utrecht University, the Netherlands

Ingrid van Dijk

Radboud University Nijmegen, the Netherlands & Lund University, Sweden

Kees Mandemakers

International Institute of Social History, Amsterdam & Erasmus University Rotterdam, the Netherlands

ABSTRACT

For the Netherlands, a rich new data source has become available which contains indexed civil certificates for multiple generations of individuals: LINKS. The current version of the dataset contains information on 1.7 million demographic events for the province of Zeeland in the 19th and early 20th centuries and will be extended to other provinces in the Netherlands in the near future. To be able to study demographic behaviour, life courses and family relations need to be reconstructed from the civil certificates. This paper describes the steps that are taken to move from the LINKS database, which contains digitised birth, marriage, and death certificates and relational information between individuals on these certificates, to LINKS-gen, which contains over six hundred thousand life courses, family reconstructions for up to seven generations, and fertility, marital, mortality, and occupational status information, ready for analysis. We present procedures for variable construction and data cleaning. Furthermore, we give a short overview of the LINKS database, discuss quality checks, and give advice on selection of relevant cases necessary to move from LINKS to LINKS-gen. The paper is accompanied by R-scripts to convert and construct the datafiles.

Keywords: Longitudinal, Longitudinal historical data, Historical micro data, Historical population database, Civil registry, Data model, LINKS, Family reconstitution, The Netherlands, Zeeland, 19th Century, 19th and 20th Centuries, Life course, HSN

e-ISSN: 2352-6343

PID article: <https://hdl.handle.net/10622/23526343-2020-0006>

PID R-scripts: <https://hdl.handle.net/10622/23526343-2020-0007>

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

© 2020, Mourits, van Dijk, Mandemakers

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

1 INTRODUCTION

In historical research, large-scale demographic databases employing automated matching procedures are becoming increasingly common (Kaplanis et al., 2017; Song & Campbell, 2017). Traditionally, these databases have been used to reconstruct families and their life courses — in so-called family reconstitutions — for the study of past demographic behaviour (Fauve-Chamoux, Bolovan, & Sogner, 2016). More recently, historical databases have also been used to study the intergenerational transmission of fertility, mortality, social status, etc. (see e.g. Bras, Van Bavel, & Mandemakers, 2013; Jennings, Sullivan, & Hacker, 2012; Knigge, 2016; Reher, Ortega, & Sanz-Gimeno, 2008; Quaranta & Sommerseth, 2018; van den Berg et al., 2019). Yet, historical databases on full populations of provinces or states are only available for a limited number of places in the world — such as Québec, Utah, and parts of Sweden — and are generally based on church records (Song & Campbell, 2017).

For the Netherlands, a rich new data source has become available: LINKS-Zeeland (Mandemakers & Laan, 2017). This dataset contains indexes of digitised birth, marriage, and death certificates from the Dutch civil registry for the full population of the province of Zeeland¹, and relational information between individuals identified on the certificates based on name-matching algorithms.² The Dutch civil registry is a high-quality historical demographic source, as from the introduction of the system the willingness to register was high, information on the certificates was highly standardised by national laws, uniformity in registration was enforced by strong procedural checks, and civil records were made in duplicate to ensure safekeeping (Vulsma, 1988). In the near future, the LINKS dataset will be expanded to include the life courses and family relations from more provinces of the Netherlands. About 700,000 birth, 200,000 marriage, and 650,000 death certificates from Zeeland are matched through indexed names and event dates. This allows for the reconstruction of (parts of) over 600,000 life courses and up to seven generations of families, starting in 1812. About 150 years of certificates have been included in the relational database: LINKS-Zeeland. To be able to study life courses or families, LINKS-Zeeland needs to be transformed from the LINKS output structure into a database suitable for research with statistical analyses: LINKS-gen.

We present how LINKS can be transformed into a 'pedigree format', which is the standard data format in the biomedical sciences for intergenerational studies (Lange et al., 2013; Purcell et al., 2007) and contains identifiers for individuals and their mother and father, as well as other demographic and relational information. The transformation of LINKS-Zeeland into one rectangular file, containing relational information and life-course observations, is a complicated endeavour. For the user, a large time investment is required to transform LINKS-Zeeland before they can perform their statistical analyses. For example, users need to decide how they deal with conflicting matches, unmatched certificates, and incomplete life course or family reconstructions. This is not only inconvenient for users of LINKS-Zeeland, but it also opens the door to variations in how users select cases and reconstruct families. Every new user will operationalise LINKS-Zeeland in a different fashion, which eventually hampers the comparability of scientific results (Ashkpour, 2019, p. 181; Garijo et al., 2014). Moreover, limitations in LINKS-Zeeland are not always immediately clear. Data in LINKS-Zeeland is of high quality, but only if the right selections are made on the data (van den Berg et al., 2020). Yet, flags for likely incomplete or complete family and life course reconstructions cannot be directly retrieved from the database, and have to be implemented as separate variables. To make LINKS-Zeeland more readily available for research, we present LINKS-gen: one rectangular data frame that is ready for statistical analyses and contains basic demographic indicators and flags for quality control. The standardisation increases the clarity and straightforwardness of LINKS-Zeeland, making it more attractive to researchers who wish to use it for research.

In this paper, we briefly describe the construction of LINKS-Zeeland and present how LINKS-Zeeland can be transformed into LINKS-gen which contains life course and family reconstructions. First, we give a short overview of the civil registry and its digitisation. Second, we briefly discuss the matching procedure that links birth, marriage, and death certificates within and between generations. This includes an explanation of how relationships between persons were established within and between generations. Third, we introduce a script that uses the established matches between certificates to

1 For more information about the history of Zeeland, see Bras (2002), Mourits and Janssens (2020,forthcoming), Priester (1998), and Wintle (1985).

2 LINKS-Zeeland 2017.01 dataset is based on the indexed civil certificates as it was archived by the WieWasWie database at December 2016. The data set was custom made for the Genes, Germs & Resources project (www.ru.nl/genesandgerms) and is also available in IDS format (Alter & Mandemakers, 2014).

reconstruct a database with each individual's life course information, relations to their parents, and spouses, as well as several other characteristics. The script that transforms LINKS-Zeeland into LINKS-gen is published together with this article.³ Fourth, we give an overview of the variables that are available in LINKS-gen. Finally, we end with some instructions on how to select cases in LINKS for research.⁴

2 LINKS: A SHORT OVERVIEW

LINKS-Zeeland refers to the matched indexes of civil certificates for the Dutch province of Zeeland. The LINKS-Zeeland 2017.01 release contains around 700,000 birth certificates (1796–1912), 200,000 marriage certificates (1796–1937), and 650,000 death certificates (1796–1962). The number of available certificates will increase over time, as new certificates become public every year.

LINKS-Zeeland is part of a larger endeavour to reconstruct the population of the entire Netherlands. The LINKing System for historical family reconstruction (LINKS) is an algorithm to match civil certificates from the Dutch civil registry. The Dutch civil registry was founded in the French period and is one of the oldest in the world. The early adaptations of the civil registry go back to July 12th 1796 for areas that were annexed by France, in the most southern part of Zeeland, Zeeuws-Vlaanderen, and in Limburg.⁵ In the rest of the country, the civil registry was introduced in 1811. However, it took until January 1st 1812 before the civil registry covered the entire country, including the unannexed part of Zeeland (Vulsma, 1988).

Civil certificates from the birth, marriage and death registers become publicly available with a delay of 100, 75 and 50 years, respectively (Burgerlijk Wetboek [Dutch civil code], article 1:17A). Regional historical archives in the Netherlands have digitised the civil registry with the help of volunteers who manually entered details from all available certificates. This effort started around 1995 and was initially led by the *National Archive* in the Hague. Nowadays, the *Dutch Family Centre* (CBG) collects the indexed and scanned certificates and subsequently makes them available to the general public through the website *WieWasWie* [WhoWasWho]. Nearly all marriage certificates from between 1796 and 1932 have been indexed and made available for public access. Digitisation of publicly accessible birth and death certificates is complete for the provinces of Drenthe, Friesland, Gelderland, Groningen, Limburg, Noord-Brabant, Utrecht, and Zeeland. The available data were cleaned by the Internationaal Instituut voor Sociale Geschiedenis (IISG) [International Institute for Social History, IISH].

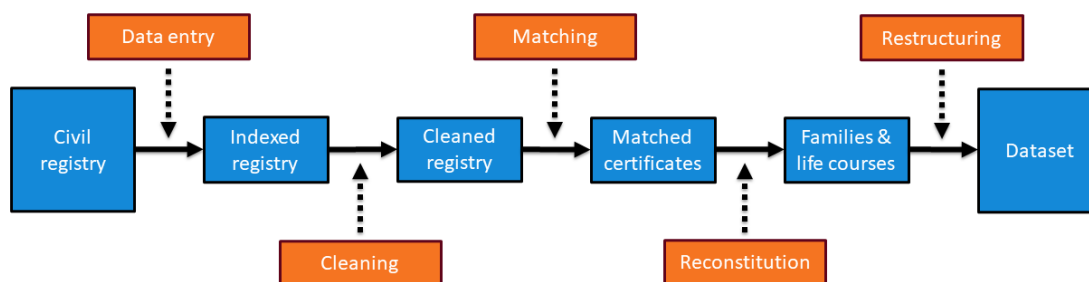
An overview of the steps between the original documents to the LINKS-gen database is shown in Figure 1. Certificates from the civil registries are indexed by volunteers from the regional archives. These indexes at least contain the names and ages of newborns, newly-weds, and the deceased, as well as the name of their parents and — if available — spouse. The names and ages of index persons, their spouse, and their parents on civil certificates are matched, so that families can be reconstructed (Mandemakers, Bloothoof, & Laan, forthcoming; Mandemakers & Laan, 2017; Raad et al., 2020; Schraagen, 2014). This provides the user with a relational database in which the relations between an individual's birth, marriage, and death certificates are used to reconstruct life courses. This very basic life course is further developed by using the relations between an individual's marriage certificate with the birth, marriage, and death certificates of his/her children, and further adding of occupational and location data (see documentation by Mandemakers and Laan (2017) for a detailed description). In this paper, we show how this information is restructured into LINKS-gen, a data set that is ready for analysis.

3 The script to build the database is also available on GitHub or the EHPS-network repository and uses R, open source software that is freely available. LINKS-gen can be exported to any other software package.

4 LINKS is available to researchers via the IISG. Please contact Kees Mandemakers (kma@iisg.nl), Auke Rijpma (a.rijpma@uu.nl), or Richard Zijdemans (r.zijdemans@iisg.nl) for access to the database, or with questions or feedback regarding LINKS-Zeeland, LINKS-gen, or the LINKS project.

5 Most of Limburg was annexed on July 12th 1796 in the Département de la Meuse-Inférieure, other parts of Limburg were incorporated in the Département de la Roër on May 1st 1798 (Vulsma, 1988).

Figure 1 Steps in building LINKS datasets for research from the civil certificates



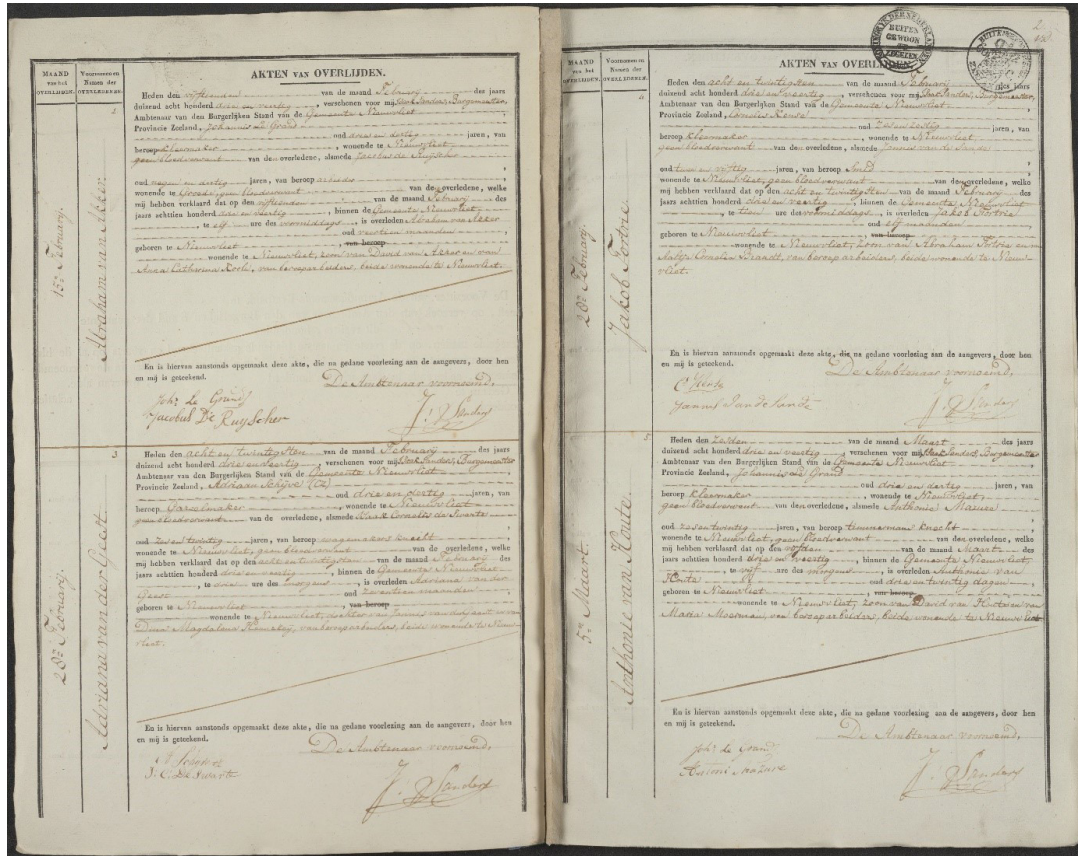
Notes: The cleaned registry, matched certificates, families & life courses, and harmonised dataset are all stored in a MySQL database and in Linked Data format. Matched certificates, families & life courses, and the harmonised dataset are also available in the format of the Intermediate Data Structure (Alter & Mandemakers, 2014). In this paper, we briefly summarise the pipeline to more exhaustively discuss the restructuring of the final dataset, which is made available as a csv-file in pedigree format.

Ever since its implementation, the Dutch civil registry has been of high quality, as standards for the municipal bureaucratic procedure were very precise. Birth, marriage, and death certificates were kept in separate books, made in duplicate, controlled by local judiciaries, and stored at separate locations (see <https://iisg.amsterdam/en/hsn/data/sources> for a short description of the administrative systems in the Netherlands or Vulsma (1988) for a more extensive overview). Initially, forms were handwritten, but between 1815 and 1850 most municipalities switched to pre-printed forms. Due to the precise administrative system, the recorded vital events are likely to represent population vital events well. However, the civil registry only includes information on family ties that (1) left a paper trail and (2) were in accordance with the law. As a result, unmarried cohabitation is not in the archives, as partners were not legally married. Furthermore, pre- and extramarital children were not matched to the father, unless the father was legally able to marry the mother — i.e. was not married himself — and acknowledged the child. These acknowledgements were added as a side note in the margin of the civil certificates. Previous studies on a sample of the Zeeland and Noord-Holland population showed that this can affect about 5% of all parents (Kok, 1991; van den Berg et al., 2020), as 2% of all parents conceived children before they were married and another 3% of all parents had children with a partner that they never married (van den Berg et al., 2020). To complicate matters further, acknowledgement of a child does not necessarily mean that the registered father is the biological father. Sometimes, upon marriage a spouse would acknowledge his wife's existing children to ensure their inheritance rights, even though there was no biological relationship. Thus, although most information contained in the civil certificates reflect the relations between the individuals quite well, imprecisions and oversights are part of the civil registry system.

Table 1 gives an overview of the information that civil servants were legally obliged to include on each civil record. Birth certificates contain the date, time, and place of birth; sex and full name of the newborn; the full name, occupation, and place of residence of the parents; and the full name, age, occupation, and place of residence of the witnesses. Marriage certificates contain the full name, age, place of birth, occupation, place of residence, and full names of former partners of both spouses; full name, age, occupation, and place of birth of the parents; full name, age, occupation, and place of residence of the witnesses; and whether children were acknowledged. Death certificates contain the full name, age, occupation, place of residence, and time of death of the deceased; full name, age, occupation, place of residence, family relation of the witnesses; and the place of birth; full name of the (deceased) spouse; as well as the full name, occupation, and place of residence of the parents.

Indexes of the civil certificates include at least the names of the subject/ego, the names of the subject's parents, subject's age, the event date, and the municipality of registration. Besides this standard information, other variables are included depending on the province and type of certificate. In Zeeland also all occupational titles of subjects and parents were included except for the birth certificates. Parents without an occupation are not always explicitly marked as having no occupation. The index of the marriage certificates also contains information on the bride's and groom's place of birth, whereas the death certificate contains both the municipality of residence and the place of death.

Figure 2 Pre-printed forms of death certificates from Nieuwvliet, 1843



Notes: Although the exact wording may change per municipality or province, texts on birth, marriage, and death certificates are standardised to a high degree. Remarks on acknowledgement of children, foundlings, multiple births, name changes, divorce dates, and other less common events are sometimes included as a side note in the margins. Side notes are generally not standardised and difficult to read due to a lack of space.

3 MATCHING PROCEDURE

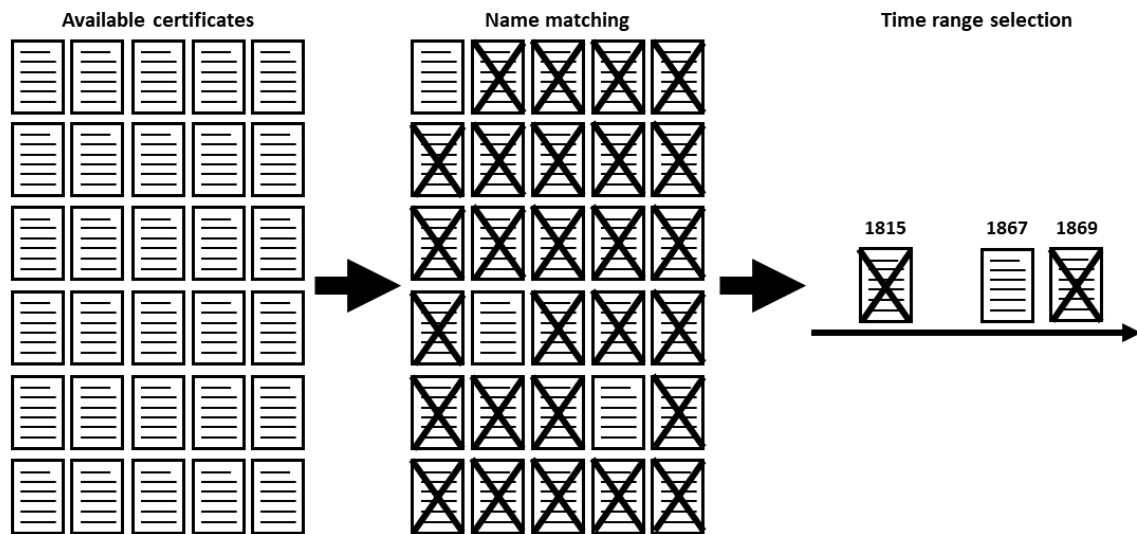
The process of matching the civil certificates of birth, marriage and death in LINKS is based on the names of subject, spouses, and parents on the civil certificates as well as time-constraints based on the dates of the events, age at marriage, and age at death (Mandemakers & Laan, 2017). An overview of the matching procedure is shown in Figure 3. The number of possible matches was limited by using the time range in which each vital event might have occurred, dramatically reducing the number of potential mismatches. Possible matches were then identified by the names of ego, father, mother, and/or spouse on the certificates whilst controlling for small errors in spelling of names in the source or during data entry, using a Levenshtein algorithm (Mandemakers & Laan, 2017; Wagner & Fisher, 1974). With the LINKS algorithm, record matching was done in a conservative manner, limiting the number of false positives as indicated by the number of non-unique matches, which indicate 'overmatching'. As such, the procedure focused on retrieving correct matches, rather than retrieving every possible one. Combined with time-constraints, 85% of all established matches between birth, marriage, and death certificates and about 75% of all established matches between parents and children were a result of exact string comparisons. The remaining matches — 15% and 25% — were retrieved using the Levenshtein algorithm with a maximum distance of 1 or 2, depending on the length of the string, for each part of the names. In LINKS-Zeeland, certificates were only matched within the province of Zeeland, so that at least part of the certificates of individuals who migrated to Zeeland or — more common — out-migrated from Zeeland to another province in the Netherlands or abroad could not be matched. Nevertheless, a sizable part of the certificates could be matched to another certificate, as shown in Table 2.

Table 1 Available information in the Dutch civil registry

General		
municipality registration date event date event time		
Birth certificate	Marriage certificate	Death certificate
<i>Informant</i> first name last name age occupation place of residence relation to the newborn	<i>Informant</i> - - - - -	<i>Informants</i> first name last name age occupation place of residence relation to deceased
<i>Child</i> first name last name sex place of birth - - -	<i>Bride & Groom</i> first name last name - place of birth age occupation name former spouse	<i>Deceased</i> first name last name sex place of birth age occupation name (former) spouse
<i>Parents</i> first name last name occupation place of residence	<i>Parents bride/groom</i> first name last name occupation place of residence	<i>Parents</i> first name last name occupation place of residence
<i>Witnesses</i> first name last name age occupation place of residence -	<i>Witnesses</i> first name last name age occupation place of residence relation to bride/groom	<i>Witnesses</i> - - - - -
<i>Other remarks</i> acknowledgement of extramarital status foundling multiple birth name changes -	<i>Other remarks</i> acknowledgement of premarital children - - - divorce date & location	<i>Other remarks</i> unsystematic - - - -

Notes: The table shows information that municipalities were legally obliged to enter according to the *Burgerlijk Wetboek* [Dutch civil code] of 1838. Women keep their family name in the civil registry after marriage. Informants are those who report the vital event to the local clerk. Informants on birth certificates can be fathers, maternal grandfathers, midwives, or local doctors. Often, this is indicative of whether the newborn is born in or out of wedlock. For an analysis, see Mandemakers (2017). Informants on death certificates are those who found the body of the deceased, usually a parent or the spouse. Witnesses are legally required attendants who testify that events were registered correctly. Remarks are generally added as side notes in the margins, but are not standardised.

Figure 3 Matching procedure in LINKS



Notes: Matches between certificates are based on two criteria: name matching and time ranges. For example, a death certificate states that Hendrik Verkleij, the son of Gerardus Verkleij and Jannetje Geelen, was 10 years old when he died on February 1st 1841. The names on the death certificate are matched to the names on birth certificates to retrieve unique matches. This leaves all certificates where the first and last names of the son, father, and mother differ less than 2 characters from: Hendrik Verkleij, Gerardus Verkleij, and Jannetje Geelen. From the remaining certificates, the certificate is selected if it fits chronologically. This person's birthday should be between February 1nd 1841 and February 2st 1842. If the age at the time of the event is unknown, logical ranges were used (see [Mandemakers & Laan, 2017](#)).

The share of matched certificates depends on the type of certificates involved. In Table 2, we show the matching rate between certificates for historical cohorts. Birth and marriage certificates had the highest matching rate, both within (own birth to marriage) as well as between generations (birth of a child to marriage parents). Within generations, birth certificates were matched to a marriage certificate in 85% of all cases, so that marriage certificates were available for 32% of all newborns. Birth and marriage certificates were matched to the marriage certificate of the parents in 83% and 79% of all cases. However, matching birth to death certificates is more difficult as death certificates are available for only 71% of all cases. In part, the relatively low matching rate between birth and death certificates was caused by out-migration. Matching death with birth certificates was more successful in terms of matched death and birth records. 80% of all Zeeland death certificates could be retraced to a birth certificate. The difference between these two indicators shows that more than 10% of matches were probably missed due to out-migration, as a sizable share of the population left the province in the 19th century ([Priester, 1998](#); [van den Berg et al., 2020](#)). Missing death certificates due to migration more strongly affected those who died at older ages than those who died early in life, as young children were less likely to leave the province ([van den Berg et al., 2020](#)).⁶ 66% of all marriage certificates were matched to a death certificate and 60% of all death certificates were matched to the marriage certificate of the parents, which is an indication that matching was more problematic for individuals who lived longer — and may have had more opportunities to leave the province — than for individuals who died as children.⁷

6 Young children could migrate with their parents. However, married couples generally migrated upon marriage and migration was much less common when married couples had small children ([Kok, 1997](#)). Therefore, information on child mortality is relatively complete, but becomes less reliable for older children.

7 Censored observations mainly occur between ages 15 and 50, as persons may migrate out of the region of observation. After age 50, migration is low, so that the number of censored cases is low ([van den Berg et al., 2020](#)).

Table 2 Percentage of successful matches between certificates, LINKS-Zeeland

	Match (A-B)	Match (B-A)	Cohort
<i>Intragenerational matches</i>			
Birth–Death	70.8%	79.9%	Born: 1812–1862
Birth–Marriage	31.9%	85.3%	Marriage: 1862–1912
Marriage–Death	65.8%	32.0%	Marriage: 1812–1862
<i>Intergenerational matches</i>			
Birth–Marriage parents	82.2%	-	Born: 1842–1912
Marriage–Marriage parents	79.4%	-	Married: 1887–1937
Death–Marriage parents	60.1%	-	Born: 1812–1862

Notes: Historical cohorts were selected. For birth and death certificates: all birth cohorts that could be followed at least 100 years. For birth and marriage certificates: all marriage cohorts with 50 years of prior observation. For marriage and death certificates: all birth cohorts that could be followed at least 100 years. For birth and parental marriage certificates: all births that had 30 years of prior observations. For marriage and parental marriage certificates: all marriages with 65 years of prior observations.

The percentage of matches was measured in two ways. Match (A-B) shows the upward percentages, whereas matched (B-A) gives backward percentages. For example, 70.8% of all men and women who were born between 1812 and 1862 had a matched death certificate. Conversely, 79.9% of all deceased persons who were born between 1812 and 1862 had a matched birth certificate.

The backward percentages (B-A) — for example from death to birth certificates — are higher than the upward percentages (A-B), because backward percentages are less affected by out-migration and, therefore, form a better indication of the quality of the matching procedure.

Matches were made in a conservative manner, preferring a high number of correct matches over a low number of missed matches, as indicated by a low number of multiple matches per certificate. Therefore, the percentage of matched cases should reflect the percentage of correct matches. Unmatched certificates can either indicate that (1) matches between available certificates were not established, or (2) that the corresponding certificate was not available due to migration.

4 FROM MATCHED CERTIFICATES TO INDIVIDUALS AND FAMILIES

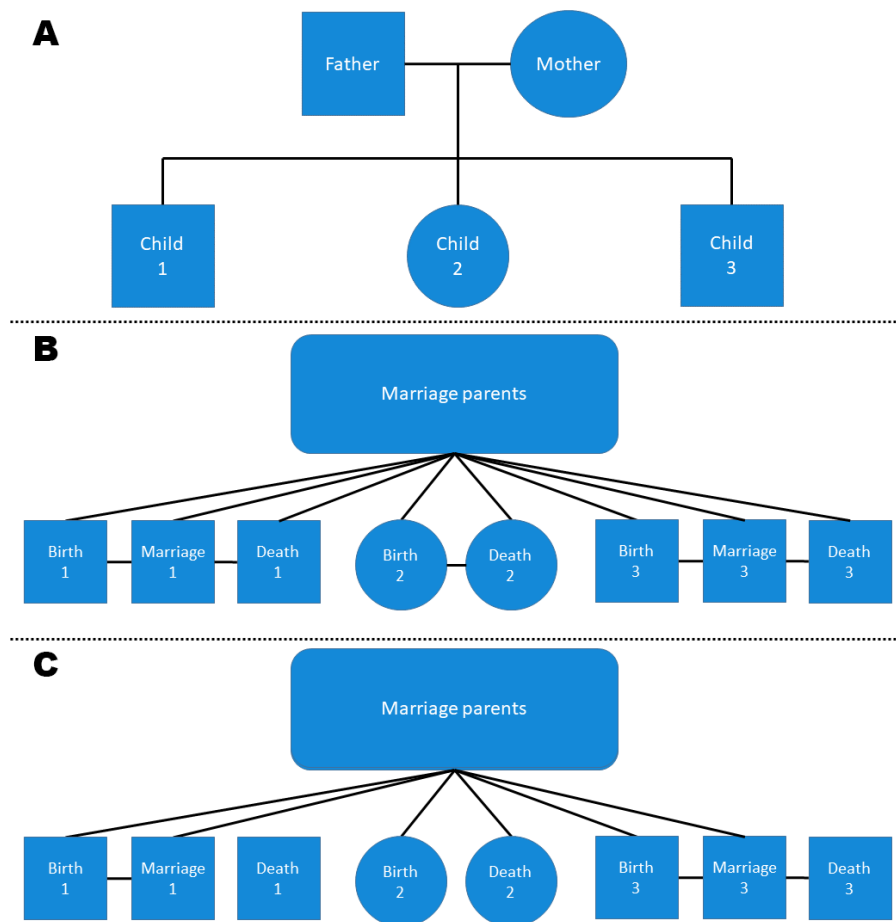
Mandemakers and Laan (2017) used the matched certificates in LINKS database to make life course and family reconstructions. Persons are defined based on the intragenerational matches between birth, marriage, and death certificates. A person identifier is assigned to each person in a cluster of matched certificates, so that records belonging to the same individual have common person identifiers. As parents also received common person identifiers, individual lives can be ordered into families over different generations. However, this procedure is not as straightforward as it might appear.

Figure 4 gives an example of how the civil certificates of three siblings are matched into life courses and how these basic life courses are ordered into sibships through the parental marriage certificate. Panel A shows an example family consisting of a mother, father, two sons, and one daughter. Panel B shows how the life courses and family relations can be deduced from the Dutch civil registry, provided that all certificates are matched. Ideally, all civil certificates are matched to the marriage certificate of the parents B as well as the other available certificates. However, this might not always be possible.

Panel C shows how a family reconstitution, that is, the reconstruction of life courses and families, works with limited information. Compared with Panel B four relations are missing: the death certificate of child 1 is not matched to any other certificate, the death certificate of child 2 matches only with the parental marriage, and the death certificate of child 3 only matches with his own marriage certificate. When certificates can still be related to another certificate of the same person, this has no effect on the

quality of the reconstructed life course or family. The death certificate of child 3 is matched through his own marriage certificate, which matches with his own birth and parental marriage certificate. As a result, his life course is not censored, nor are observations on child 3 scattered. However, life course and family reconstructions are more problematic when a certificate cannot be related to another certificate of the same person. The death certificate of child 2 was not matched to her birth certificate. As a consequence, her life course is censored and another person with only a known age of death is created. Child 2 now appears twice in the reconstructed family, so that one might think that the family consists of 4 rather than 3 children. In LINKS-gen we restructure the data to deal with this problem. Finally, the death certificate of child 1 can neither be matched to his own marriage or birth certificate, nor to his parents' marriage certificate. Therefore, the life course of the child 1 is censored. Moreover, a new family is created that exists of one child with known mortality information, unknown parents, and no birth information. This loose certificate can only be excluded from analyses by making selections on the data.

Figure 4 Family relations in a pedigree format and in LINKS



Notes: Panel A shows a family with two parents and three children: father, mother, an eldest son, a middle daughter, and a youngest son. Panel B shows how the life courses and family relations can be deduced from the Dutch civil registry, provided that all matches between certificates succeed. Panel C shows how a family reconstitution works with limited information.

In Panel C, the death certificate of the oldest son (child 1) cannot be matched to the family, which results in a new line in the LINKS database with an unmatched death certificate. The birth and death certificate of the daughter (child 2) cannot be matched directly, because the name of the child was identical to a sibling's name or too similar to a twin's name. In this case, post-hoc matches are made within families based on the names of the parents and time range. The death certificate of the youngest son (child 3) is matched through his marriage certificate, because a match with the parental marriage certificate was not established.

The output from the life course and family reconstructions in LINKS-Zeeland 2017.01 is produced from the data in LINKS database. All established matches between certificates are stored with extra variables that indicate the quality of the match, for example, the number of person names on which a match is based. Currently, the release consists of three tables: BIRTH_DEATH, MARRIAGE_LINES, and PERSONS. BIRTH_DEATH contains the established matches between birth and death certificates as well as matches to the marriage certificates of the parents. A record consists of either matched births and deaths, or unlinked birth and death certificates. The table MARRIAGE_LINES contains matches to the bride's and groom's certificates in BIRTH_DEATH as well as matches to their parents' marriage certificates. Multiple occurrences of the same person were assigned the same identifier, so that each record in the two tables refers to a unique 'person', who has been assigned its own unique identifier and is listed in the table PERSONS.

5 RESTRUCTURING THE DATA

At this point, the information that is delivered to a user of LINKS-Zeeland needs to undergo several transformations to be prepared for data analysis. Selections on the data, post-hoc matching, and knowledge of the data are required before the reconstructed life courses and families can be analysed (van den Berg et al., 2020). Information in LINKS-gen is structured according to the pedigree format (Lange et al., 2013; Purcell et al., 2007) of which an example is shown in Table 3. In a pedigree file, each row contains information on a person including relations to their father, mother, and pedigree, meaning all related individuals. This information makes it very easy to study inheritance patterns, familial clustering, and intergenerational transmission of fertility, mortality, social status, etc.

Table 3 *Example of the pedigree file structure*

Id pedigree	Id person	Id father	Id mother	Sex	Variables
1	1	0	0	Female	
1	2	0	0	Male	
1	3	2	1	Male	
1	4	2	1	Female	
1	5	2	1	Male	
1	6	0	0	Female	
1	7	5	6	Female	
1	8	5	6	Female	
2	9	0	0	Female	
2	10	0	0	Male	
2	11	10	9	Female	
3	12	0	0	Male	
4	13	0	0	Female	
4	14	0	0	Male	
4	15	14	13	Male	
4	16	0	0	Female	
4	17	14	16	Female	

Notes: The table shows family structures for four different pedigrees. Pedigree 1 consists of three generations: parents (person 1–2), children (person 3–5), the partner of person 5 (person 6), and grandchildren (person 7–8). Pedigree 2 consists of a family with parents (person 9–10) and one daughter (person 11). Pedigree 3 consists of a man (12) whose parents are unknown. Pedigree 4 consists of a father (person 14) who married two different women (persons 13 and 16). Together with his first wife he had a son (person 15) and with his second wife he had a daughter (person 17).

The size of a pedigree is only limited by the available kin relations. When the grandchildren, children, or unmatched individuals from pedigrees 1–4 get children of their own, these children are added to the pedigree of their parents. If this causes pedigrees to overlap, these two separate pedigrees should

be merged into one pedigree, for example when person 8 from pedigree 1 has children with person 12 from pedigree 3.

In earlier steps, relations between parents and children were defined. However, the data on individual characteristics needs to be restructured to make the database ready for analysis. There are several issues that need to be solved: the information on life courses and relatives is divided over several tables; incomplete birth data information needs to be supplemented with estimations from marriage/death dates and ages at marriage/death; missed matches within families need to be restored; and variables need to be constructed that flag from which certificates information is derived. The steps that we present below provide the user with LINKS-gen.

Birth dates were entered as available in the database. But, when persons had no matched birth certificate, the first and latest possible birthdate were estimated using the age at marriage and/or the age at death. In this case, the birthdate was logically deduced by calculating the mean of the estimated birth range. If both an estimation from marriage and death certificates are available, information from both certificates is used to narrow down the period in which a person was born. To the birth information, we added marital information. Since persons may marry more than one time, we reserved five series of fields for marriage information, which is the maximum number of marriages in the dataset, and also constructed a variable on the number of marriages by counting the number of marriages certificates., Thereupon, marriages were put in chronological order and the dates of marriage were included.

Death certificates also include newborns who were dead upon registration and for whom no separate birth certificates were registered. These children were flagged as 'dead before registration'. In the current database, we use the date of death as the date of birth, as these were not registered separately. The logically deduced date of birth is a reasonable estimate for those children which died before they could be administered a birth certificate. However, the correct birthdate may differ up to three days from the registered death date for children who were born over the weekend. This problem is exacerbated for children who were born before a holiday as the difference between their birth date and registration date increased, increasing the likelihood that they were not stillborn but died before they were registered with the municipality, due to the delay in registration. Earlier work has estimated that most of the Dutch children who were dead upon registration were stillborn, but that about one in three died before their registration with the municipality (van Poppel, 2018; van Poppel, Bijwaard, Ekamper, & Mandemakers, 2012).

Information originating from certificates of marriage or death could not always be matched to a birth certificate. The person could have been born in another province, could have been born before the start of the birth registration, or after the period within which birth certificates become available. However, in specific cases there were also other reasons for missed matches. In over 10,000 cases, parental names and time ranges matched, but not the names of individuals. Matches were missed for twin siblings with too similar names — for example, 'Gerritje' and 'Gerrit' or 'Johan' and 'Joran'. The same problem can occur for children that received the same first name as an earlier deceased sibling. In both cases, observations on the same individual were falsely recorded as observations on different individuals within the same family. To solve this problem, we applied a post-hoc matching procedure. Within families, we added unmatched death certificates to birth and marriage certificates of children without a death certificate, provided that the estimated birth date overlapped with the actual date of the birth certificate, and that an unmatched death certificates was relatable to only one child in the family. This provided us with 9,580 (2.9%) extra matched death certificates out of 331,298 unmatched death certificates in the case of LINKS-Zeeland.

Finally, we constructed variables on whether individuals had a twin sibling and on the date of last observation. In case no death certificate was matched, dates of marriage and childrens dates of birth were used to indicate the last date at which an individual was observed alive. For women, when applicable, the date of last childbirth was used as the last observation date. For men, we used the date of last childbirth minus nine months, as not all fathers may have lived until childbirth. In the data set we report the date and age of last observation as well as whether the observation was a birth, first marriage, second marriage, the birth/conception of a child, or death.

6 INFORMATION IN LINKS-GEN

The restructured and recoded family reconstitutions result in LINKS-gen, which structures the life course and family reconstructions in a pedigree format (Lange et al., 2013; Purcell et al., 2007). In LINKS-gen multiple generations of intertwined families are recorded by using the matches between the individual birth, marriage, and death certificates to the marriage certificates of one's parents. Each row contains information on an individual with a unique identifier: *Id_person*. This person can be directly related to his father, mother, and spouse(s) through the corresponding identifiers: *Id_father*, *Id_mother* and *Id_partner_1* to *_5*. Up to five marriage partners are available, as no individual had more than 5 known partners. Relations to siblings or children are identified indirectly. Sibling sets can be reconstructed by selecting individuals with a common father and mother to retrieve full siblings or with only one common father or mother to retrieve half siblings. Children are related to their parents with the identifiers: *Id_father* and *Id_mother*. It should be noted that these sibship and child sets can be quite large as Zeeland was a high fertility region. As a result, families with over 15 children are not unheard of.

Table 4 and Figure 5 show which information is available. For each person, we indicate the sex as available on the original certificate, and constructed variables on whether or not a person had a twin sibling, the date of and age at last observation, and the total number of known marriages. We also flag whether birth, marriage, and death certificates were available for the subject, his/her father, and his/her mother. Birth information includes the birthdate, municipality of birth, and age of the parents at birth. If no birth certificate was known, the value of *B_date*, *B_min_date*, and *B_max_date* was logically deduced as explained in section 5. Death information includes age at death, date of death, municipality of death, and flags on whether a newborn died before he could be administered a birth certificate. Marital information shows the number of marriages as well as the age at marriage, municipality of marriage, and the identifier of the spouse for each marriage.

LINKS-gen does not contain explicit pedigree identifiers. As we use general population data, many men and women appear in more than one pedigree. Standard practice dictates that these pedigrees ought to be merged together (see e.g. Sinnwell, Therneau, & Schaid, 2014). However, such a strategy is not advisable, because LINKS would fall apart in a few, large and non-informative, pedigrees. Rather, pedigrees should be based on the selection of families with certain characteristics. Thereupon, a random sample of families ought to be drawn to prevent overrepresentation of large pedigrees or individuals who reproduced with multiple spouses (see e.g. van den Berg et al., 2019). The current format allows for efficient sampling of individuals and families, so that users can select cases with relevant information and study tailor-made pedigrees.

Occupational information is stored in a separate file, the contents of which are shown in Table 5 and Figure 6. In the civil registry, occupations were only recorded in concordance with vital events. In the case of LINKS-Zeeland, occupational information was generally available at an individual's own marriage and death certificate. Marriage certificates seem to be the best source for female occupations (Boter & Woltjer, 2020; Walhout & van Poppel, 2003). Furthermore, occupational titles of the parents are also available on the marriage and death certificates of children. On birth certificates the parental occupations are only available for the island of Walcheren. The occupational database gathers all these occupations and lists them by the combination of *Id_person* and *Date*. The file further lists the calculated age at observation, as well as the certificate containing the occupational information. Occupational information is enriched by recoding occupations into HISCO (van Leeuwen, Maas & Miles, 2002; 2004), HISCLASS (van Leeuwen & Maas, 2011), HISCAM (Lambert, Zijdemans, van Leeuwen, Maas, & Prandy, 2013), and SOCPO (Van De Putte & Miles, 2005) by way of the HSN-HISCO occupational title release (Mandemakers et al., 2018). Further, the occupational codings are converted from HISCO into OCC1950 (U.S. Bureau of the Census, 1950) using the HISCO-OCC1950 conversion table (Mourits, 2017), and converted from OCC1950 into Nam-Powers-Boss (Nam & Boyd, 2004; Nam & Powers, 1983), OCCSCORE (IPUMS, 2017b), PRESGL (Siegel, 1971), and SEI (Duncan, 1961) using conversion tables from IPUMS (2017). The latter systems are useful for comparisons with US data. OCC1950 and HISCLASS are almost identical ($\rho: 0.97$), whereas the correlation between HISCAM and Nam-Power-Boss, OCCSCORE, PRESGL, or SEI is moderate to high ($\rho: \approx 0.65$) (Mourits, 2019, p. 83).

Table 4 Overview of available variables

	Variable	Description	Type
Identifiers	Id_person	Person identifier (primary key)	-
	Id_mother	Person identifier mother	-
	Id_father	Person identifier father	-
	Id_partner_1 to _5	Person identifier partner	-
General	Sex	Sex (m/f/u)	unaltered
	Twin	Has a twin brother or sister (y/n)	constructed
Last Entry	LastEntryDate	Date of last observation	constructed
	LastEntryAge	Age at last observation	constructed
	LastEntryCert	Certificate containing last observation	constructed
Certificate	B	Birth certificate is available (y/n)	flag
	M	Marriage certificate is available (y/n)	flag
	D	Death certificate is available (y/n)	flag
	Postlink_D	Death certificate matched on parental names only	flag
	M_parents	Parent's marriage certificate is available (y/n)	flag
	B_mother	Mother's birth certificate is available (y/n)	flag
	B_father	Father's birth certificate is available (y/n)	flag
	D_mother	Mother's death certificate is available (y/n)	flag
	D_father	Father's death certificate is available (y/n)	flag
Birth	B_date	Date of birth	unaltered
	B_max_date	Latest possible date of birth	logical edit
	B_min_date	Earliest possible date of birth	logical edit
	B_location	Municipality of birth	unaltered
	B_age_mother	Age father at birth	constructed
	B_age_father	Age mother at birth	constructed
	Death	D_date	Date of death
D_age		Age at death	unaltered
D_location		Municipality of residence at time of death	unaltered
D_deadonregistration		Dead at declaration of birth	unaltered
Marriage	Marriage_N	Total number of marriages	constructed
	M_age_1 to _5	Age at marriage	unaltered
	M_date_1 to _5	Date of marriage	unaltered
	M_location_1 to _5	Municipality of marriage	unaltered

Notes: Each row contains a person. Unknown sex generally refers to children who died before registration and received no name. Flags indicate from which certificates the available information is retrieved. Unaltered information is transcribed from a civil certificate, logical edits are deduced from transcribed data, and constructed variables combine information from multiple certificates.

Figure 5 Example of the LINKS-gen pedigree table

	Id_person	Id_mother	Id_father	Sex	Twin	LastEntryDate	LastEntryAge	LastEntryCert	B	M	D	postlink_D	M_parents
1	37810246	37810248	37810247	m	0	1873-09-22	78.00030	Death	0	1	1	0	0
2	37356391	37346768	37346767	m	0	1862-03-04	69.53194	Death	0	1	1	0	0
3	37309757	37309759	37309750	m	0	1883-12-09	01.51303	Death	0	1	1	0	0
4	37643415	38132159	38132156	m	0	1903-04-16	72.75172	Death	1	1	1	0	1
5	37451462	37402351	37402350	m	0	1874-07-26	87.99914	Death	0	1	1	0	0
6	37845264	37552794	37552793	m	0	1879-11-19	67.64279	Death	1	1	1	0	0
7	37400376	37665234	37719621	m	0	1900-11-29	61.96710	Death	1	1	1	0	1
8	37410633	37410635	37410634	m	0	1882-05-17	70.19178	Death	1	1	1	0	0
9	37782929	37649713	37649712	m	0	1869-06-20	70.27392	Death	0	1	1	0	0
10	37748735	37563527	37563526	m	0	1859-06-08	57.41943	Death	1	1	1	0	0
11	37833211	37977622	38463888	m	0	1902-03-20	82.28512	Death	1	1	1	0	1
12	37480114	37520049	37520046	f	0	1878-12-02	73.62238	Death	1	1	1	0	1
13	37590051	37379888	37379887	m	0	1864-10-13	65.14308	Second marriage	0	1	0	0	0

Table 5 Overview of available SES variables

	Variable	Description	Type
Identifiers	Id_person	Unique person identifier	-
General	Date	Date at which the occupation was registered	unaltered
	Age	Age at which the occupation was registered	logical edit
	Location	Municipality where occupation was registered	unaltered
Certificate	Cert	Type and sequence number of the certificate mentioning the occupation: <ul style="list-style-type: none"> • Marriage_N (marriage certificate number N) • B_N (birth certificate child number N) • M_N (marriage certificate child number N) • D_N (death certificate child number N) • Death (death certificate) 	flag
Occupation	Occupation	Occupation (standardised)	unaltered
Social status	HISCO	HISCO classification	enriched
	HISCLASS	HISCLASS score	enriched
	SOCPO	SOCPO score	enriched
	HISCAM	HISCAM score	enriched
	OCC1950	1950 US Census Bureau classification	enriched
	OCCSCORE	Estimated occupational income	enriched
	PRESGL	PRESGL score	enriched
	SEI	SEI score	enriched
NPBOSS	Nam-Powers-Boss score	enriched	

Notes: Rows refer to occupational entries. Each record is identified by the combination of Id_person & Date. Flags indicate from which certificates is retrieved. Unaltered information is transcribed from a civil certificate, logical edits are deduced from transcribed data, constructed variables combine information from multiple certificates, and enriched cases use data from external matching table on social status. HISCO, HISCAM, HISCLASS, and SOCPO codings are retrieved from Mandemakers et al. (2018), OCC1950 codings from Mourits (2017) and OCCSCORE, PRESGL, SEI, and NPBOSS from IPUMS (2017).

Figure 6 Example of the LINKS-gen occupational table

	Id_person	Date	Age	Location	Location_ACODE	Location_CBS	Occupation	HISCO
1	35240955	1821-04-21	NA	Hulst	11408	677	winkelier	41030
2	35240969	1852-11-09	31.12729038	Aardenburg	11020	648	werkman	99900
3	35240976	1814-08-16	NA	Arnemuiden	11419	649	visser	64100
4	35240979	1913-06-25	NA	Kruiningen	10077	685	rijksontvanger	31020
5	35240980	1913-06-25	NA	Kruiningen	10077	685	zonder beroep	Without occupation / unemployed
6	35240982	1886-08-29	NA	Middelburg	10122	687	bierbrouwer	77810
7	35240983	1886-08-29	NA	Middelburg	10122	687	zonder beroep	Without occupation / unemployed
8	35240996	1889-04-24	18.01270502	Zaamslag	11173	731	zonder beroep	Without occupation / unemployed

7 USING THE DATABASE

A comparison of LINKS-Zeeland with sampled life courses from Zeeland has shown that LINKS-Zeeland contains high quality data. Missed and false matches between certificates had little effect on the quality of life course reconstructions. However, reconstructed life courses are only of sound quality if individuals stayed in Zeeland their entire lives, because certificates from other regions have not been indexed yet (van den Berg et al., 2020). Therefore, the database is only suited for research if certain selections on the data are made to prevent biased descriptive or inferential statistics, such as Cox proportional hazard models, ordinary least squares regressions, and maximum likelihood logistic regressions (Alter, Devos, & Kvatko, 2009; Gill, 1997). LINKS-gen has included flags that indicate whether birth, marriage, and death certificates were available for an individual or his/her parents.

LINKS-Zeeland can only produce descriptive statistics for specific subgroups of the population. Unlike census or population register data, the civil registry does not indicate how many people were living somewhere at any point in time. Individuals were only included in the civil registry when they experienced a vital event, so that observations on places of residence are scattered and migration is only measured when individuals had children or died in another town. As a result, it is next to impossible to reliably estimate populations at risk, even when the full civil registry for the Netherlands is available. However, such estimates can be given for subgroups for which the population at risk is known or can logically be deduced. For example, life expectancies cannot be calculated, as we do not know the population at risk between ages 0–100. However, mortality before age 5 and after age 50 can be studied, as parents with small children and older individuals were unlikely to migrate (Kok, 1997). Similarly, fertility rates cannot be estimated for the entire population, but many related statistics can be calculated for parental couples that married and died in Zeeland, as these likely resided in the same province for their entire life.

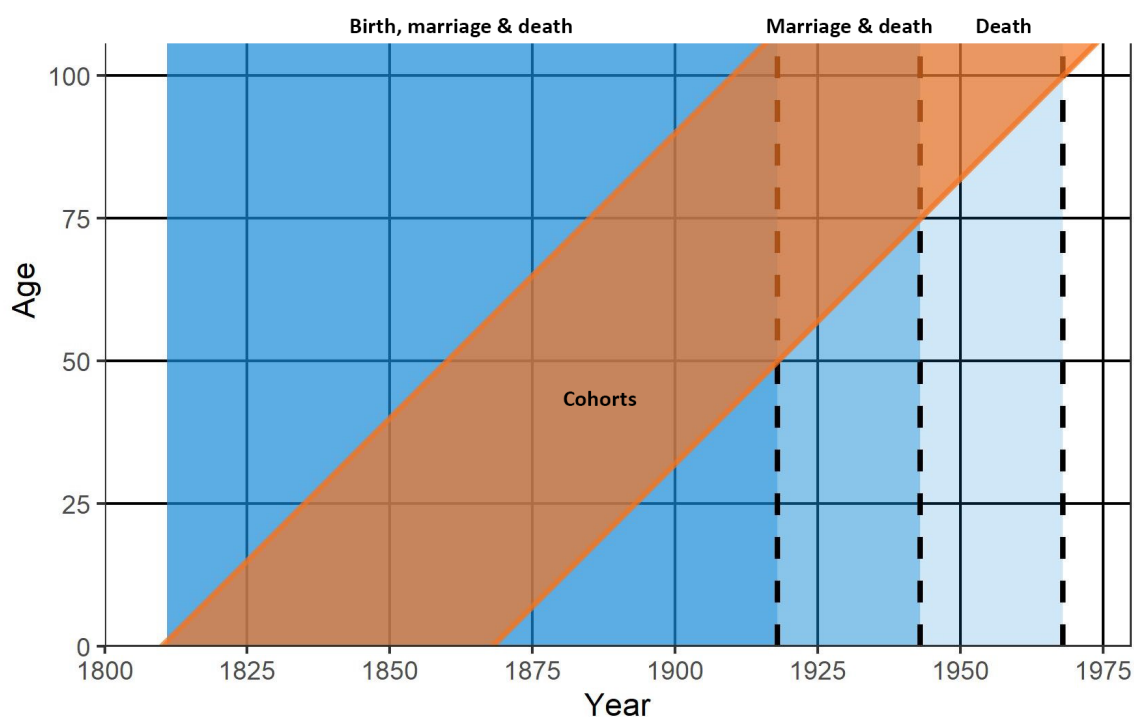
Flags in LINKS-gen indicate the quality of reconstructed life courses. For research on fertility behaviour, migration histories, and other individual characteristics full observation of individuals is necessary, as censored observations produce informed missings. The best indicator for a full observation is the flagged availability of a death certificate, as it indicates that an individual died in Zeeland, which makes it more likely that he also experienced other vital events in the region. However, some return migration is observed at higher ages (van den Berg et al., 2020). Furthermore, the availability of a death certificate does not mean that information is available on children. For women, information on their children is nearly always available, as a mother's name automatically appears on a birth certificate. However, the father's name was only known when he was married to the mother or not married to another women. Therefore, a flagged marriage certificate needs to be available to study male fertility behaviour. In other words, researchers should check the B, M, and D flags to check whether the reconstructed life courses are not censored.

Similarly, flags can also be used to indicate the quality of family reconstructions. For family reconstructions, the marriage certificate of the parents is the most important certificate. Children are matched to their parents' marriage certificate. For an individual, the availability of a parental marriage certificate signals that information on the preceding generations and parental life course might be available. Provided at least one parental death certificate is flagged, intergenerational effects of parental fertility behaviour, migration, or other parental characteristics on children can be studied. Furthermore, the availability

of a parental marriage and death certificate usually indicates that the reconstruction of a sibship is of sound quality (van den Berg et al., 2020). This means that inferences on parental birth intervals and the total number of siblings are likely based on complete observations. Moreover, if at least one sibling has a flagged death certificate life course information for this sibling will be relatively complete as well, allowing for the study of familial clustering of events such as mortality and fertility as well as socioeconomic status. Hence, the flags *M_parents*, *D_father*, and *D_mother* indicate whether family reconstructions are uncensored, while *D* flags indicate whether a child's life course can be used to study of intergenerational effects and sibling similarity.

We further advice users of LINKS-gen to consider the window of observation carefully. For the entire country, birth, marriage, and death certificates are of sound quality from 1812. More recent certificates are protected by privacy laws, in the current dataset from the year 1912, 1937, and 1962 for births, marriages and deaths, respectively. Between 1796 and 1811, certificates are also available for regions within Zeeland and Limburg, but the quality of the administration appears to be less. Furthermore, the civil administration started in 1811, but seemed to be of better quality from 1812 onwards. To give a few examples of how this influences research on LINKS-Zeeland: families can only be reconstructed if they were started after 1811, as otherwise one or more siblings may be missing. Infant mortality can only be studied between 1812 and 1912, as birth certificates are no longer available after 1912. Two-generation reconstructions of life courses are only available for the 1812–1862 marital cohorts, as their children are born between 1812 and 1892, and their grandchildren between approximately 1830 and 1912, after which birth certificates are no longer available. The Lexis diagram in Figure 7 gives a systematic overview of the availability of certificates over time.

Figure 7 Lexis diagram of available information in LINKS-gen



8 CONCLUSION

To clear the way for social research on historical life courses, we presented a way to convert LINKS into one comprehensible dataset that can readily be analysed. The script that transforms LINKS-Zeeland into LINKS-gen is published together with this article. Information from birth, marriage, and death certificates is ordered into one logical data frame. This structure will be kept in place for future releases of LINKS, and is designed to simplify and standardise the use of the database. The restructured dataset

provides users with three advantages. First, users can more efficiently invest their time in understanding and operationalizing the dataset. Second, researchers can use a standardised version of LINKS, which increases the comparability and replicability of future research. Finally, users can now easily select 'good cases' for their research, as we flag cases that are likely to be relatively complete. Instructions for the selections for different types of demographic research can be found in this paper. Together, these factors make LINKS-gen attractive to researchers who wish to engage in social, economic and demographic historical research on the 19th-century Netherlands.

The LINKS database contains an enormous amount of detailed information on life courses and family relations from 19th century the Netherlands. It gives a rare insight into family relations in the 19th century, as historical databases on full populations of provinces or states are only available for a limited number of places in the world, such as Québec and Utah (Song & Campbell, 2017). In the upcoming years, LINKS will be developed further until it spans all provinces of the Netherlands. The resulting database will be an impetus to future historical, demographic, social, and biomedical research, as it provides new opportunities for research on kin networks, intergenerational similarities in demographic behaviour, and spatial disparities (see e.g. Knigge, 2016; van Dijk & Mandemakers, 2018).

Another strong benefit of a digitised civil registry is that it can be matched with other sources and serve as a basis for further historical reconstructions. Currently, an infrastructure is being developed by the CLARIAH project to reconstruct the socioeconomic context of the 19th century, both on the individual and contextual level (Hoekstra et al., 2018). LINKS serves as the backbone for this project, as it clearly structures the (partial) life courses of everyone who lived in 19th century the Netherlands. Demographic information from LINKS-Netherlands can be matched with other sources, such as inheritance tax registers (Peeters, de Vicq de Cumplich, & Gelderblom, 2019) to get a grasp of wealth and possessions, or causes of death registers (Janssens, 2019) to model changes in the disease environment to study human stature. Moreover, once information from the population registers becomes structurally available, information on coresidency and migration can be added. Hence, information on individual wealth, mortality, stature, cohabitation, and migration behaviour will be contextualised into intergenerational, familial networks.

9 AVAILABILITY LINKS

LINKS is available to researchers via the IISG. Please contact Kees Mandemakers (kma@iisg.nl), Auke Rijpma (a.rijpma@uu.nl) or Richard Zijdemans (r.zijdemans@iisg.nl) for access to the database, or with questions or feedback regarding LINKS-Zeeland, LINKS-gen, or the LINKS project in general.

ACKNOWLEDGEMENTS

The authors would like to thank all volunteers, Gerrit Bloothoof, Jan Kok, Fons Laan, the Genes, Germs & Resources project, as well as CLARIAH WP4 for their ongoing work on the LINKS database.

REFERENCES

- Alter, G., Devos, I., & Kvetko, A. (2009). Completing life histories with imputed exit dates: A method for historical data from passive registration systems. *Population*, 64(2), 293–318. doi: [10.3917/popu.902.0327](https://doi.org/10.3917/popu.902.0327)
- 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>

- Ashkpour, A. (2019). *Theory and practice of historical census data harmonization: The Dutch historical census use case: A flexible, structured and accountable approach using linked data technology* (PhD thesis Erasmus University Rotterdam). Retrieved from <http://hdl.handle.net/1765/114388>
- Boter, C., & Woltjer, P. (2020). The impact of sectoral shifts on Dutch unmarried women's labor force participation, 1812–1929. *European Review of Economic History*, 00, 1–35. doi: [10.1093/ereh/hez020](https://doi.org/10.1093/ereh/hez020)
- Bras, H. (2002). *Zeeuwse meiden: Dienen in de levensloop van vrouwen, ca. 1850–1950* (PhD thesis Utrecht University). Amsterdam: Aksant.
- Bras, H., Van Bavel, J., & Mandemakers, K. (2013). Unraveling the intergenerational transmission of fertility: genetic and shared-environment effects during the demographic transition in the Netherlands, 1810–1910. *The History of the Family*, 18(2), 116–134. doi: [10.1080/1081602X.2013.803491](https://doi.org/10.1080/1081602X.2013.803491)
- Burgerlijk Wetboek [Dutch civil code] (1837). 's-Gravenhage: Ter Algemeene Lands Drukkerij. Retrieved from <https://catalog.hathitrust.org/Record/006923362>
- Burgerlijk Wetboek [Dutch civil code] (2019–01–01). Retrieved 10 June 2019 from <https://wetten.overheid.nl/BWBR0002656/>
- Duncan, O. D. (1961). A socioeconomic index for all occupations. In A. Reiss et al. (Eds.), *Occupations and social status*. Glencoe, Illinois: Free Press of Glencoe.
- Fauve-Chamoux, A., Bolovan, I., & Sogner, S. (2016). A global history of historical demography: Half a century of interdisciplinarity. Bern: Peter Lange.
- Garijo, D., Alper, P., Belhajjame, K., Corcho, O., Gil, Y., & Goble, C. (2014). Common motifs in scientific workflows: An empirical analysis. *Future Generation Computer Science*, 36, 338–351. doi: [10.1016/j.future.2013.09.018](https://doi.org/10.1016/j.future.2013.09.018)
- Gill, R. D. (1997). Nonparametric estimation under censoring and passive registration, *Statistica Neerlandica*, 51(1), 35–54. Available from doi: [10.1111/1467-9574.00036](https://doi.org/10.1111/1467-9574.00036)
- Hoekstra, R., Meroño-Peñuela, A., Rijpma, A., Zijdemans, R., Ashkpour, A., Dentler, K., Zandhuis, I., & Rietveld, L. (2018). The dataLegend ecosystem for historical statistics. *Journal of Web Semantics*, 50, 49–61. doi: [10.1016/j.websem.2018.03.001](https://doi.org/10.1016/j.websem.2018.03.001)
- IPUMS. (2017a). NPBOSS50: Nam-Powers-Boyd occupational status score, 1950 basis. Retrieved from https://usa.ipums.org/usa-action/variables/NPBOSS50#description_section
- IPUMS. (2017b). OCCSCORE: Occupational income score. Retrieved from https://usa.ipums.org/usa-action/variables/OCCSCORE#description_section
- IPUMS. (2017c). PRESGL: Occupational prestige index, Siegel. Retrieved from https://usa.ipums.org/usa-action/variables/PRESGL#description_section
- IPUMS. (2017d). SEI: Duncan socioeconomic score. Retrieved from https://usa.ipums.org/usa-action/variables/SEI#description_section
- Janssens, A. (2019). The Amsterdam health & disease database. Retrieved from <https://www.ru.nl/rich/our-research/research-groups/radboud-group-for-historical-demography-and-family/completed-research-projects/completed-research-projects/amsterdam-health-disease-database-1854-1940/>
- Jennings, J. A., Sullivan, A. R., & Hacker, J. D. (2012). Intergenerational transmission of reproductive behaviour during the demographic transition. *Journal of Interdisciplinary History*, 42(4), 543–569. Available from doi: [10.1162/JINH_a_00304](https://doi.org/10.1162/JINH_a_00304)
- Kaplanis, J., Gordon, A., Shor, T., Weissbrod, O., Geiger, D., Wahl, M., Gershtovits, M., Markus, B., Sheikh, M., Gymrek, M., Bhatia, G., MacArthur, D. G., Price, A. L., & Erlich, Y. (2018). Quantitative analysis of population-scale family trees with millions of relatives. *Science*, 360(6385), 171–175. doi: [10.1126/science.aam9309](https://doi.org/10.1126/science.aam9309)
- Knigge, A. (2016). Beyond the parental generation: The influence of grandfathers and great-grandfathers on status attainment, *Demography*, 53(4), 1219–1444. doi: [10.1007/s13524-016-0486-6](https://doi.org/10.1007/s13524-016-0486-6)
- Kok, J. (1991). *Langs verboden wegen: De achtergronden bij buitenechtelijke geboorten in Noord-Holland*. Hilversum: Uitgeverij Verloren.
- Kok, J. (1997). Youth labor migration and its family setting, the Netherlands 1850–1940. *History of the Family*, 2(4), 507–526. doi: [10.1016/S1081-602X\(97\)90027-0](https://doi.org/10.1016/S1081-602X(97)90027-0)
- Lambert, P. S., Zijdemans, R. L., van Leeuwen, M. H. D., Maas, I., & Prandy, K. (2013). The construction of HISCAM: A stratification scale based on social interactions for historical comparative research. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 46(2), 77–89. doi: [10.1080/01615440.2012.715569](https://doi.org/10.1080/01615440.2012.715569)

- Lange, K., Papp, J. C. Sinsheimer, J. S., Sripracha, R., Zhou, H., & Sobel, E. M. (2013). Mendel: The Swiss army knife of genetic analysis programs, *Bioinformatics*, 29(12), 1568–1570. doi: [10.1093/bioinformatics/btt187](https://doi.org/10.1093/bioinformatics/btt187)
- Mandemakers, K. (2017). En hoe zit het eigenlijk met de getuigen? Professionele en relationele achtergrond van de getuigen in de geboorteakten van Nederland, 1812–1922. In Th. Engelen & T. Riswick (Eds.), *Van historische informatiekunde naar historische sensatie. Libretto voor Onno Boonstra* (pp. 61–78). Nijmegen: Valkhof Pers.
- Mandemakers, K., Bloothoof, G., & Laan, F. (forthcoming). LINKS: The LINKing System for historical family reconstruction in the Netherlands.
- Mandemakers, K., & Kok, J. (2020). Dutch lives: The Historical Sample of the Netherlands 1987-): Development and research. *Historical Life Course Studies, Online first*, 1–47. Retrieved from <http://hdl.handle.net/10622/23526343-2020-0001?locatt=view:master>
- Mandemakers, K., & Laan, F. (2017). LINKS dataset Genes Germs and Resources, WieWasWie Zeeland, Civil Certificates, version 2017.01 [Data file and code book]. Amsterdam: IISH.
- Mandemakers, K., Mourits, R. J., Muurling, S., Boter, C., van Dijk, I. K., Maas, I., Van De Putte, B., Zijdemans, R. L., Lambert, P., van Leeuwen, M. H. D., van Poppel, F., & Miles, A. (2018). *HSN-standardized, HISCO-coded and classified occupational titles, release 2018.01*. Amsterdam: IISH.
- Mourits, R. J. (2017). HISCO-OCC1950 crosswalk. DANS. doi: [10.17026/dans-zap-qxmc](https://doi.org/10.17026/dans-zap-qxmc)
- Mourits, R. J. (2019). *Exceptional lives, extraordinary families: Familial clustering of longevity in the 19th and early-20th centuries*. Enschede: Ipskamp.
- Mourits, R. J. & Janssens, A. (2020). Spatial clustering of longevity in Zeeland, 1812–1962: How stable, behavior-associated environmental characteristics explain the local clustering of longevity. *Annales de démographie historique*. Forthcoming.
- Nam, C. B., & Boyd, M. (2004). Occupational status in 2000: Over a century of census-based measurement. *Population Research and Policy Review*, 23(4), 327–358. doi: [10.1023/B:POPU.0000040045.51228.34](https://doi.org/10.1023/B:POPU.0000040045.51228.34)
- Nam, C. B., & Powers, M. G. (1983). *The socioeconomic approach to status measurement: With a guide to occupational and socioeconomic status scores*. Houston, Texas: Cap & Gown Press.
- Peeters, R., de Vicq de Cumplich, A., Jonkers, J., & Gelderblom, O. (2019). Tafel-Vbis: Summary of all death duties in the Netherlands in 1921. DANS. doi: [10.24416/uu01-qg9q8b](https://doi.org/10.24416/uu01-qg9q8b)
- Priester, P. (1998). *Geschiedenis van de Zeeuwse landbouw circa 1600–1910*. Wageningen: Landbouwniversiteit Wageningen.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3), 559–575. doi: [10.1086/519795](https://doi.org/10.1086/519795)
- Quanjer, B., & Kok, J. (2020). Drafting the Dutch: Selecting biases in Dutch conscript records in the second half of the 19th century. *Social Science History*, 44(3), 501–525. doi: [10.1017/ssh.2020.13](https://doi.org/10.1017/ssh.2020.13)
- Quaranta, L., & Sommerseth, H. L. (2018). Introduction: Intergenerational transmissions of infant mortality using the Intermediate Data Structure (IDS), *Historical Life Course Studies*, 7(2), 1–10. Retrieved from <http://hdl.handle.net/10622/23526343-2018-0014?locatt=view:master>
- Raad, J., Mourits, R. J., Rijpma, A., Schalk, R., Zijdemans, R., Mandemakers, K., & Meroño-Peñuela, A. (2020). Linking Dutch civil certificates. *3rd Workshop on Humanities in the Semantic Web (WHiSe) conference proceedings*. Heraklion, Greece. Retrieved from http://whise.cc/2020/papers/WHiSe_2020_paper_8.pdf
- Reher, D. S., Ortega, J. A., & Sanz-Gimeno, A. (2008). Intergenerational transmission of reproductive traits in Spain during the demographic transition. *Human Nature*, 19(1), 23–43. doi: [10.1007/s12110-008-9032-6](https://doi.org/10.1007/s12110-008-9032-6)
- Schraagen, M. (2014). Aspects of Record Linkage (PhD thesis Leiden University). Retrieved from <http://hdl.handle.net/1887/29716>
- Siegel, P. M. (1971). *Prestige in the American occupational structure*. Chicago: University of Chicago.
- Sinnwell, J. P., Therneau, T. M., & Schaid, D. J. (2014). The kinship2 R package for pedigree data. *Human Heredity*, 78(2), 78–93. doi: [10.1159/000363105](https://doi.org/10.1159/000363105)
- Song, X., & Campbell, C. (2017). Genealogical microdata and their significance for social science. *Annual Review of Sociology*, 43, 75–99. doi: [10.1146/annurev-soc-073014-112157](https://doi.org/10.1146/annurev-soc-073014-112157)
- U.S. Bureau of the Census. (1950). *Alphabetical index of occupations and industries: 1950*. Washington, D.C.: U.S. Bureau of the Census. Retrieved from <https://usa.ipums.org/usa/resources/volii/Occupations1950.pdf>

- Van De Putte, B., & Miles, A. (2005). A social classification scheme for historical occupational data. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 38(2), 61–94. doi: [10.3200/HMTS.38.2.61-94](https://doi.org/10.3200/HMTS.38.2.61-94)
- van den Berg, N., van Dijk, I. K., Mourits, R. J., Janssens, A., Slagboom, P. E., Janssens, A. A. P. O., & Mandemakers, K., (2020). Families in comparison: An individual-level comparison of life course and family reconstructions between population and vital event registers, *Population Studies*, 1–20. doi: [10.1080/00324728.2020.1718186](https://doi.org/10.1080/00324728.2020.1718186)
- van den Berg, N., Rodríguez-Girondo, M., van Dijk, I. K., Mourits, R. J., Mandemakers, K., Janssens, A. A. P. O., Beekman, M., Smith, K. R., & Slagboom, P. E. (2019). Longevity as top 10% survivors and beyond is transmitted as a quantitative trait. *Nature Communications*, 10, 1–12. doi: [10.1038/s41467-018-07925-0](https://doi.org/10.1038/s41467-018-07925-0)
- van Dijk, I. K., & Mandemakers, K. (2018). Like mother, like daughter: Intergenerational transmission of infant mortality and clustering in Zeeland, the Netherlands, 1833–1912. *Historical Life Course Studies*, 7(2), 28–46. Retrieved from <http://hdl.handle.net/10622/23526343-2018-0003?locatt=view:master>
- van Leeuwen, M. H. D., Maas, I., & Miles, A. (2002). *HISCO: Historical International Standard Classification of Occupations*. Leuven: Leuven University Press.
- van Leeuwen, M. H. D., Maas, I., & Miles, A. (2004). Creating a historical international standard classification of occupations: An exercise in multinational interdisciplinary cooperation. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 37(4), 186–197. doi: [10.3200/HMTS.37.4.186-197](https://doi.org/10.3200/HMTS.37.4.186-197)
- van Leeuwen, M. H. D., & Maas, I. (2011). *HISCLASS: A historical social class scheme*. Leuven: Leuven University Press.
- van Poppel, F. (2018). Legal rules and practices of registration and burial of stillbirths in 19th and 20th century Netherlands. In C. Rollet, G. Clavandier, N. Sage Pranchère, P. Charrier, & V. Gourdon (Eds.), *Morts avant de naître: La mort périnatale*. Tours: Presses Universitaires François-Rabelais.
- van Poppel, F., Bijwaard, G. E., Ekamper, P., & Mandemakers, K. (2012). Historical trends in the correlation of sibling deaths in infancy in the Netherlands. *Biodemography and Social Biology*, 58(2), 87–115. doi: [10.1080/19485565.2012.720448](https://doi.org/10.1080/19485565.2012.720448)
- Vulsma, 1988. *Burgerlijke stand en bevolkingsregister*. 's-Gravenhage: Centraal Bureau voor Genealogie.
- Wagner, R. A., & Fischer, M. J. (1974). The string-to-string correction problem. *Journal of the Association for Computing Machinery*, 21(1), 168–173. Available from doi: [10.1145/321796.321811](https://doi.org/10.1145/321796.321811)
- Walhout, E., & van Poppel, F. (2003). De vermelding des beroeps: eene ijdele formaliteit? Twee eeuwen vrouwelijke arbeidsmarkt in Nederlandse huwelijksakten. *Tijdschrift voor Sociale Geschiedenis*, 29(3), 301–332.
- Wintle, M. (1985). Aspects of religion and society in the province of Zeeland (Netherlands) in the nineteenth century (Unpublished PhD thesis University of Hull).