Family matters: a genealogical inquiry into the familial component of longevity

Berg, N.M.A. van den

CHAPTER 3

FAMILIES IN COMPARISON:

AN INDIVIDUAL-LEVEL COMPARISON OF LIFE COURSE AND FAMILY RECONSTRUCTIONS BETWEEN POPULATION AND VITAL EVENT REGISTERS

Niels van den Berg[1,2*]
Ingrid K. van Dijk[2*]
Rick J. Mourits[2*]
Angelique A.P.O. Janssens[2]
P. Eline Slagboom[1,3]
Kees Mandemakers[4]

[1] Department of Biomedical Data Sciences, section of Molecular Epidemiology, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, the Netherlands
[2] Radboud Group for Historical Demography and Family History, Radboud University, Erasmusplein 1, 6525 HT Nijmegen, the Netherlands
[3] Max Planck Institute for Biology of Ageing, Joseph-Stelzmann-Str. 9B, D-50931 Cologne, Germany
[4] International Institute of Social History, Cruquiusweg 31, 1019 AT Amsterdam

*. Shared first authorship

Population Studies (March 2018)

Abstract

It remains unknown how different types of sources affect the reconstruction of life courses and families in large-scale databases, increasingly common in demographic research. Here, we compare the family and life-course reconstructions for 495 individuals, simultaneously present in two well-known Dutch datasets. LINKS-Zeeland is based on a province's full-population vital-event registration data (passive registration). The HSN is based on a national sample of birth certificates, with follow-up of individuals in population registers (active registration). We compare indicators of fertility, marriage, mortality, and occupational status and conclude that reconstructions in the HSN and LINKS reflect each other well, although LINKS provides more complete information on siblings and parents, whereas the HSN provides more complete life-course information. We conclude that life-course and family reconstructions based on linked, passive registration of individuals constitute a reliable alternative to such reconstructions based on active registration, if case selection is carefully considered.

Introduction

Demographic research is increasingly conducted on large-scale longitudinal datasets. Underlying these databases are sources such as population registers, parish registers, registrations of vital events, censuses, and genealogical databases. Names, ages, birthplaces, and other personal characteristics in these sources are used to link life-course events, e.g. marriage or migration, to individuals (life-course reconstruction), and individuals to each other into family networks (family reconstruction) . Characteristics of underlying data sources may affect the completeness and quality of life-course and family reconstructions in databases[1–3]. This is particularly the case for comparisons between databases derived from active registration where individuals are followed continuously over time, registering specific events as they happen, and databases produced from passive registration where individuals are only observed when specific events, such as a birth or marriage, are registered and the separate documents are linked together[4,5]; see *Supplementary Table 1* for an overview of active and passive registration). Well-known examples of databases based on active registration include the Roteman database for Stockholm in Sweden and the Scanian Demographic Database (SEDD). Databases based on passive registration include the Utah Population Database, the English Family Reconstitutions, and Knodels German village family reconstitutions.

Databases based on passive registration can more easily miss a vital event, such as the birth of a child. Migration movements are not registered, making it unclear whether, where, and when an individual experienced vital events in another region. As households or individuals are followed actively during their lives in active registration, observations generally contain relatively complete information on individuals and their families. Out-migration is commonly observed, so that when individuals leave the municipality or the region of residence, they can easily be traced to their new place of residence. Thus, both differences in source material as well as strategies for following individuals across data sources are likely to be crucial for the quality of reconstructed lives and families in historical databases. The extent to which they result in differently reconstructed life courses and families remains unexplored in the literature, however, due to a lack of data which enable cross-checks of the same life courses and families using different sources, with the exception of Wisselgren et al. (2014)[6].

In this paper, we show a comparison of life-course and family reconstructions for the same individuals in demographic datasets derived from two different, independent data sources: one based on the Dutch population registers, reflecting active registration, and one based on Dutch vital-event registers, reflecting passive registration. Our purpose is twofold: to investigate to what extent life courses and family reconstructions are represented similarly in databases which are based on active versus passive registration, and to determine the suitability of the types of data for different research questions, including questions on life spans and mortality, marriage behaviour, and fertility. The results are of interest to researchers working with individual level longitudinal demographic data of either sort.

[1] Combined, life-course and family reconstructions form the basis of the practice known as family reconstitution. Family reconstitution is the process of reconstructing historical data on family membership and the events occuring to these family members during the course of their lives[47].

An overview of the literature

Earlier research focusing on the quality of individual-level, large-scale longitudinal demographic databases has used a variety of approaches which consider the characteristics of the source material and the logic of the construction of the database[7–10]. Other studies have used approaches based on internal consistency of databases and comparisons to external data sources such as mortality statistics in life tables. Schellekens & Van Poppel (2016) compared population-register-data to national statistics and reported that in the Historical Sample of the Netherlands (HSN) cohort life expectancy at age 30 may be overestimated for men, but not for women[11]. Adams et al. (2002) concluded that observations on migration in vital event registrations reflected migration information in population registers well[12].

One of the main drawbacks of data based on passive registrations is that analysis is usually restricted to the residentially stable part of the population and excludes those without an age at out-migration or age at death, raising issues of representativeness[4,5,13,14]. Importantly, Ruggles (1992) observed that migration causes underestimation of population-level demographic indicators, e.g. age at marriage, age at first and last birth, and number of children[15]. After migration, migrants are right-censored and demographic events are no longer observed, causing an underestimation of the number of events as well as the mean age at the corresponding events, and this is all the more problematic when the date of migration is not recorded so that only the last recorded observation can be used. If the last observation is not a death, a potential source of bias is introduced, because individuals are still at risk of experiencing events after their last observation in the population. Statistical inferences have been developed to estimate dates of last observation when censoring occurred[4,5,16]. However, there may be true differences between the migrating and non-migrating part of the population[13,17].

While approaches based on external data sources are useful instruments to judge the quality of databases, they only provide insight into deviations at the aggregate level, such as differences in mortality rates. Whether individual life courses and families are reconstructed accurately remains an open question. Some efforts have been made in this direction, as linkage success and percentages of correct matches across sources - such as parish records and census material - have been used as an indicator of data quality[6,18,19]. In addition, several studies explored the success of linking strategies by comparisons between databases (see, for instance, Wisselgren et al. (2014) for comparisons between Swedish censuses and parish registers and Massey (2017) for historical US data)[6,20]. Ruggles et al. (2018) emphasize that most studies focus on missed links (type II errors), so that false links (type I errors) are given too little attention[21]. Both errors may introduce bias into the life-course and family reconstructions. However, missed and false links not only affect whether individuals are included in demographic databases, but also whether the correct children, spouses and parents are linked to them. By paying proper attention to false links, life-course transitions may be more accurately incorporated in databases. False matches and failed matches mostly occur in sources based on passive registration where individuals are not continuously followed over time. However, direct comparisons with sources based on active registration may reveal areas where passive registration may provide more complete data.

Data

In the Netherlands, a unique opportunity has opened up to compare the individual life-course and family reconstructions in two different types of datasets. The first is based on a sample of birth certificates (HSN) and contains active registration on households originating from the nationwide population registers. The second is based on the civil registry of Zeeland (LINKS) and contains linked civil certificates of birth, marriage and death. Individuals born in Zeeland who were included in the HSN can be identified in LINKS through an identifying combination of the municipality, year and sequence number provided on each civil certificate.

The civil registry and LINKS

The civil registry

The Dutch civil registry is one of the oldest in the world, and has covered the entire country from 1812 onward. Birth, marriage, and death certificates were kept in separate books, made in duplicate, controlled by local judiciaries, and stored at separate locations (see Vulsma, 1988[22]). The Dutch civil registry of birth, marriages and deaths is a good source for life-course and family reconstructions. All certificates contain the date of the event, the date of the registration (birth and death certificates), the place of registration, the name and age of the person reporting the event, and the names and places of residence of the witnesses. The birth certificates contain - if known - the name of the father as well as the name of the mother, and the name and sex of the child. The marriage certificates contain the age, occupation, civil status before the marriage and place of residence of both spouses, the names of the bride's and groom's parents, and - if they were alive - their age, place of residence and occupation. For death certificates, one of the two informers - one informer after 1935 - reporting the death is often a spouse or parent; they report the name, occupation, age, place of residence of the person reporting and the person reporting and the deceased person. The civil registers of births, marriages and deaths become public with delays of 100, 75, and 50 years, respectively (Burgerlijk Wetboek [Dutch civil code], article 1:17A).

The LINKing System for historical family reconstruction (LINKS)

LINKS (LINKing System for historical family reconstruction) is based on digitised certificates from the civil registries, as indexed by the WieWasWie project, to reconstruct families. The Zeeland 2017.01 release of the database contains around 700 thousand birth certificates, 200 thousand marriage certificates, and 650 thousand death certificates. Multigenerational families were built using linked marriage certificates, reconstructing life courses and families (see *Figure 1*). Of the births detailed in LINKS, 81% were linked to the marriage of their parents. In total, the dataset contains almost 2 million persons covering a maximum of seven generations. Individual life-course reconstructions resulted from linking civil birth, marriage and death certificates: 68% of all birth certificates and 66% of all marriage certificates were linked to a death certificate[19]. The scope of the database is large regarding intergenerational networks of family members[23], but the successful reconstruction of life courses and families depends on the linkage of passively registered data sources. In addition, LINKS does not contain information on addresses, co-residence of individuals, migration movements, and religion.

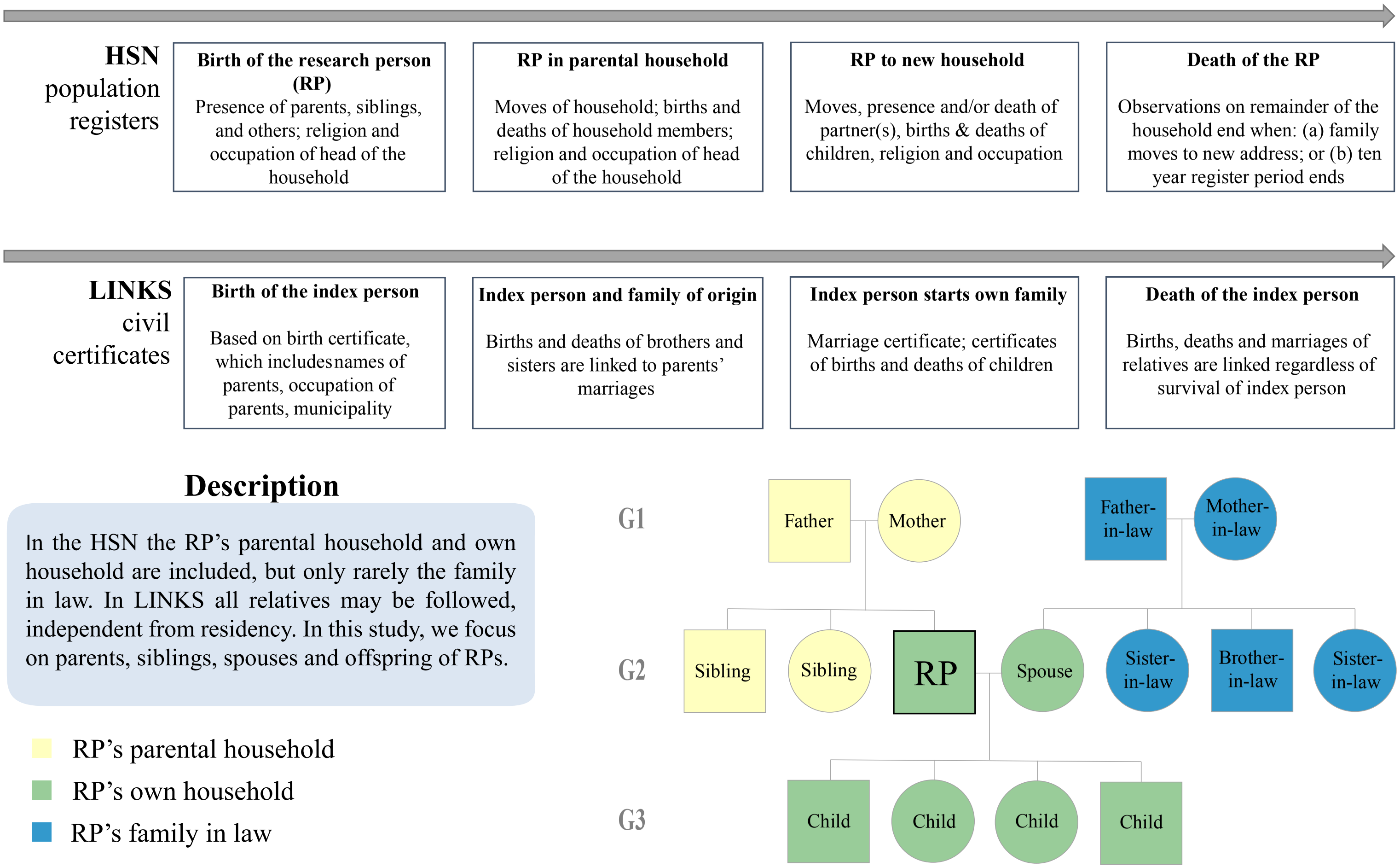


Figure 1: Data structure for HSN and LINKS.

Both LINKS and HSN can be used for life course (top two rows) and pedigree reconstruction. The top row shows the information based on the life course reconstruction in the HSN. The second row shows the information based on the family reconstruction in LINKS. The pedigree shows in general the available persons in LINKS and HSN. The green/light grey and yellow/white color represent the selection for the paper which resembles the pedigree structure that can be derived from HSN. Color descriptions first state the color version color and subsequently state the black and white version color (color/black and white).

Indexes of civil birth, marriage and death certificates were linked together, using combinations of at least two pairs of names of individuals, spouses and parents, combined with time-constraints based on age. Variations in the spelling of names, name changes, and the non-uniqueness of many names renders family reconstruction a complicated task. To prevent missed matches due to spelling variations, all first and last names were corrected for minor known variations in spelling. All name combinations for at least two persons, the individual and one or two of his or her parents and possibly a spouse or child, were matched. In the data release used here, certificates were only linked within the province of Zeeland, so that certificates of individuals who out-migrated from Zeeland to another province in the Netherlands or abroad were missed. This concerned a sizable part of the population, for example, those who migrated to Belgium and Rotterdam[24].

Population registers and the HSN

Population registers

Population registers were introduced in the Netherlands from 1850. The population registers were maintained by each municipality in large books, organized by streets or neighbourhoods. This makes it possible to follow households - and the persons in them - over time. For each household, the registers contain information on the family name, given names, sex, marital status, birth dates, death dates, birth places, address, professions, and religious denominations. For married couples, the head of the household was the male spouse. After his death, his widow would become the head of the household until her death, until remarriage, or until she moved into a household with an existing male head[25,26]. Relationships between the members of the household are included from the perspective of the head of the household, allowing the reconstruction of relationships between other household members. Movements into and out of the household as well as births and deaths were actively tracked. The books containing the population registers were replaced every ten years and updated with a coinciding nationwide census. This active registration allows the follow-up of households for longer periods of time.

In the period of research of this paper, two important changes in the population registries were implemented. In 1920 – and earlier in the large cities – the population registration was no longer ordered by street or neighbourhood, but by individual household in a card system with separate documents. From 1939 onward, the registration was no longer focused on households, but on individuals by means of personal cards. Later, in 1994, this personal card system was completely digitized. Nowadays, the system is known as the Personal Records Database (Dutch: Basisregistratie Personen, BRP) and is maintained on the national level. One year after a person's death a summary of personal and family information becomes available for scientific and genealogical research (CBG, 2019) and for specific research purposes a request can be made to Dutch government to directly access the BRP.

The Historical Sample of the Netherlands (HSN)

In the HSN, the life courses of a representative group of individuals in the Netherlands are followed. The HSN enables research on detailed life courses of individuals from the 19th century for The Netherlands[9,27]. The HSN is based on a sample of birth certificates of all individuals living in the Netherlands, stratified by cohorts of ten years for the period between 1811 and 1922 and according to regional levels of population density. The sample consists of 0.75% of the births for the period

1812 – 1872 and 0.5% of the births for the period 1873 – 1922. In total, the sample consists of about 85,500 individuals[28]. Up to now, about 40,000 of these 85,500 persons have been followed in the population registers throughout their life course. In the HSN these persons are referred to as "Research Persons" (RPs). The population-register information in the HSN is supplemented with information from the Dutch birth and marriage certificates.

In the HSN release 2010.01, entries in population registers and on personal cards of 37,137 RPs were made available[9]. For some regions, including Zeeland, the HSN starts already in 1850, when the population registers were introduced. The database includes information about the RPs' household, including co-residents, occupation, and religion. Households were, in principle, only followed as long as the RP was present in that household. Siblings and other kin were eventually lost from observation when the RP moved out of the household or died, after a follow-up to the end of the 10-year population register period. For the period after the implementation of family cards for individual households, the remaining family members were followed for up to 40 or 50 years.

Structural differences between the HSN and LINKS

Because of the sampling procedure and independent sources of information, structural differences exist between the databases in terms of the life course and family reconstructions (see *Figure 1 and Table 1*). In the HSN, siblings and parent information was only available to the extent to which family members cohabited with RPs. Therefore, questions with topics such as intergenerational and horizontal kin relations – for instance, sibling similarities in mortality – cannot be answered. Second, in LINKS individuals were observed when vital events occurred to them, their spouses, or their children. Consequently, the HSN is primarily focused on life course reconstruction and less on family reconstructions, whereas the opposite applies to LINKS, in the sense that observations on life events are used to trace family members. We will explore to what extent events of fertility marriage, migration, mortality, and occupational careers were observed and differed between the HSN and LINKS.

First, in contrast to the HSN, LINKS does not encompass unmarried cohabitation or extramarital children, which may lead to an underestimation of the number of children or siblings. Second, the lack of continuous follow-up of individuals in the civil registry makes it necessary to link certificates. The automatic record linking procedure might occasionally miss matches between vital event certificates. Moreover, certificates were only linked within the geographic area of a province, so persons were lost if they migrated to another province or country. Thus, mortality in early life was most likely measured quite accurately, but certificates of deaths and marriages occurring later in life are more likely to be unavailable. Finally, key indicators such as occupation and place of residence were only observed in concordance with vital events of individuals, their spouses or their children.

Table 1: Expected availability of demographic indicators in the HSN and LINKS

	HSN		LINKS	
	Availability on data sources	**Reason**	**Availability on data sources**	**Reason**
Parents Marriages	*Incomplete*	Not included if parents were not in household; marriage date of parents often not known	*Incomplete*	Not available for RP's who moved out of Zeeland
RP's Sibship size	*Incomplete*	Not included if siblings died before follow-up of the RP, or were born after RP moved out of the household	*Incomplete*	Not available for RP's who moved out of Zeeland
Marriages of RP	*Incomplete*	Marriages incompletely registered in population registers	*Incomplete*	Not available for RP's who moved out of Zeeland
Fertility	*Incomplete*	Offspring not included if they died before registration; no stillbirths recorded	*Incomplete*	Not available for RP's who moved out of Zeeland
Family relations	*Not always clear*	Relations within household need to be logically reconstructed for the period 1850-1862; family relations in 3rd or 4th degree may be unclear in subsequent registers	Clear	-
Occupation	Complete	Updated regularly	*Incomplete*	Not available for RPs who moved out of Zeeland; only known when a vital event was registered; measured relatively early in the life course
Later-life mortality	Complete	-	*Incomplete*	Not available for RP's who moved out of Zeeland
Extramarital fertility	Complete	Premarital fertility included; RPs who lived together but were not married	*No information*	No information on extramarital fertility
Migration	Complete	Continuous follow-up of migration in the Netherlands	*Incomplete*	Only known when a vital event was registered; persons are followed through Zeeland only
Children Child mortality	*Incomplete*	No information on offspring outside the RP's household	*Incomplete*	Not available for RP's who moved out of Zeeland

Migration patterns and occupational careers can be reconstructed from an individual's civil certificates, as well as from their children's civil certificates. Death certificates contain occupational information if the deceased person had an occupation at the time of death. Hence, more observations on occupation and place of residence were available for RP's who married or had children. Moreover, most of these vital events occur relatively early in life, so that later changes in place of residence and occupation could easily be missed. For unmarried individuals, only vital events in the family of origin and one's death certificate will be observed.

In the HSN, there were no systematic observations of events before the sampled RP was followed. Observations on RPs does not always start at birth, leading to gaps in life course and family reconstructions. The implication is that siblings who reside elsewhere or died young may not be included in the register in which the RP first appears. As a result, the count of all known siblings reflects the count of surviving siblings – the net fertility – rather than the count of all siblings ever born – the total fertility. At the same time, RPs children were identified very accurately in the HSN because RPs were, in principle followed for their entire life course. This is illustrated by[29], who showed for Tilburg (1849-99) that 99.8% of the children found in the birth registers were identified in the population registers. At the same time, stillbirths and children who died very soon after birth were usually not included in the birth or population registers, but only in the death registers (hereinafter, "lifeless reported infants"). These characteristics limit opportunities for research on events early in the life course – such as exposure to sibling mortality or the length of birth intervals – and research on intergenerational relations in longevity, mortality, and fertility.

Data construction and approach

For the comparison between the HSN and LINKS, we used persons identified in both databases who were born between 1863 and 1872 . Drawing on data from LINKS 2017.01[30] and the HSN 2010.01 population register release[31–33], the 495 Zeeland-born individuals included in the HSN were traced in LINKS via unique identifiers of the birth certificates. We analysed differences in life-course and family reconstructions of RPs in the estimation of key demographic and socioeconomic indicators. We test whether the characteristics of the databases might have lead to an underestimation in the number of links. Demographic linking strategies tend to go for precision (few false matches) at the expense of recall (few missed matches)[6,34]. Moreover, biases in the registration procedure leads to omissions in the data. Therefore, different observations between the HSN and LINKS are most likely indicative of false negatives, i.e. missed observations.

An overview of all available information in both datasets and expected completeness is provided in *Table 1*. For our analyses, we used the following indicators: sex, start and end dates of observation (HSN) or first and last observation (LINKS), birth year, and death year. We counted the number of siblings and children known, and the birth order of the RP in their family of origin. With regard to the number of siblings and children, stillbirths and infants lifeless upon civil registration were excluded, as they were unavailable in the HSN. In addition, we measured ages at first and last childbirth. Furthermore, we noted whether RPs married or not and had children or not; calculated their age at first marriage and at death; traced whether they migrated within Zeeland, outside Zeeland (HSN), or never; and tested their socioeconomic position on consistency between both datasets using HISCLASS, a social class scheme to classify historical professions[35,36].

Results

Table 2 presents the number of RPs for whom parents, siblings and children could be identified. Because entire households of individuals are actively registered in the source material underlying the HSN and observations are available for the entire country, information on parents, spouses, and children is more often available in the HSN than in LINKS, which is based on linked civil certificates from Zeeland. In the HSN, 96% of the RPs had available parent information, for a total of 932 parents. In LINKS, 82% of the RPs had available parent information (814 parents). In the HSN 1,060 children were identified (for 40% of the RPs), whereas in LINKS 810 children were identified (for 31% of the RPs). However, fewer siblings are known in the HSN than in LINKS (1,447 and 2,804 siblings, for 72 and 83% of RPs, respectively), as these were only observed if they lived together with the RP in a household. Fewer spouses were known in the HSN than in LINKS, because marriages were registered in the civil records in the first place, and may not always have been registered correctly in the population registers. A total of 233 spouses were found in the HSN (28% of the RPs), while 188 spouses were identified in LINKS (36% of the RPs). Hence, active registration increases the number of RPs with known family relations, but might be related to missed events that occurred outside of an RPs household or in other registers. This difference between events within and outside the household does not exist for passively registered sources. For both datasets, the number of individuals without spouses and children appears to be high. However, many individuals in Zeeland did not reach reproductive ages, as infant and child mortality in Zeeland was very high, reaching up to 50% in some municipalities and years[37,38].

Table 2: Available family ties in the HSN and LINKS for the selected 495 RPs from the 1862-1871 Zeeland cohort

Relatives	Sample size	RPs with known relatives (%)
HSN		
Parents	*932*	475 (96)
RPs	495	
Siblings	1447	336 (72)
Spouses	233	138 (28)*
Children	1060	196 (40)
LINKS		
Parents	814	407 (82)
RPs	495	
Siblings	2804	413 (83)
Spouses	188	177 (36)
Children	810	151 (31)

The 233 spouses are identified using the population registers. These numbers are used in the paper itself. When adding marriage certificates to the population registers, we identified 237 spouses and 277 unmarried RPs. Combined, the population registers and marriage certificates identify 324 spouses and 270 unmarried RPs. The RPs with known relatives refer to the number of RPs with for example known parents (N=475). Spouses are based on the number of marriages. Hence, one RP could have had multiple spouses.

Comparisons between the databases were conducted in two ways. First, we compared all individuals for whom relevant observations can be expected in both databases separately, with the purpose of exploring all life course and family reconstructions *(Table 3A)*. Because the mean scores in this table were based on different RPs, these means must be interpreted for each dataset separately. Second, we analysed only the subsets for which we could reconstruct life courses in an identical way, hence, we selected individuals for whom a relevant observation may be expected in both databases *(Table 3B)*. Both tables show key demographic information for all RPs for whom it is possible to know whether they experienced the demographic event. Cases without information on the relevant selection criteria were not included. Differences between the HSN and LINKS in demographic indicators in *Table 3B* indicate differences in the reconstructions of life courses and families between the HSN and LINKS, whereas differences in these indicators in *Table 3A* may also be caused by differences between the subsets of individuals for whom information was available.

Table 3: Number of available cases and mean score on demographic indicators after selection of best cases in the HSN and LINKS separately (3A) and simultaneously (3B)

	Date selection	Indicators	Table 3A				Table 3B		
			HSN N	LINKS n	HSN N (mean)	LINKS N (mean)	HSN-LINKS N	HSN N (mean)	LINKS N (mean)
Siblings	Lived until age 30	Number of RPs with identified siblings (mean number of siblings)	221	-	221 (3.9)	-	186	186 (3.8)	-
	Lived until age 30	Birth order	221	-	221 (1.8)	-	186	-	186 (6.6)
	Parents known	Number of RPs with identified siblings (mean number of siblings)	-	407	-	407 (6.7)	186	186 (1.8)	-
	Parents known	Birth order	-	407	-	407 (4.2)	186	-	186 (4.0)
Marriages	Lived until age 30	Ever-married*	221	146	122 (55.2)a	124 (84.9)	138	62 (44.9)a	118 (85.5)
	Lived until age 30	Never-married*	221	146	99 (44.8)b	22 (15.1)	138	776 (55.1)b	20 (14.5)
	Ever-married	Age at first marriage	137	178	137 (28.4)c	178 (26.3)	97	97 (27.7)c	97 (26.8)c*
Children	Ever-married	No children identified*	122	178	13 (9.5)d	26 (14.6)	97	8 (12.9)d	15 (15.5)d*
	Children identified	Age at first childbirth	196	152	196 (27.0)	152 (26.5)	146	146 (26.8)	146 (26.6)
	Children identified	Age at last childbirth	196	152	196 (37.4)	152 (36.6)	146	146 (37.4)	146 (36.8)
	Children identified	Number of RPs with identified children (mean number of children)	196	152	196 (5.4)	152 (5.2)	146	146 (5.4)	146 (5.4)
Migration behavior	Lived until age 18	Never moved*	236	157	62 (26.3)	90 (57.3)	149	56 (37.6)	86 (57.7)
	Lived until age 18	Moved within Zeeland*	236	157	78 (33.1)	67 (42.7)	149	71 (47.7)	63 (42.3)
	Lived until age 18	Moved out of Zeeland*	236	157	95 (40.3)	-	149	21 (14.1)	-
Age at death	All	All ages at death	409	313	409 (40.8)	313 (34.7)	306	306 (33.6)	306 (33.9)
	Lived until age 18	Ages at death after 18	236	157	236 (69.4)	157 (67.4)	149	149 (67.1)	149 (67.1)
	Lived until age 50	Ages at death after 50	204	134	204 (75.1)	134 (73.6)	126	126 (73.6)	126 (73.6)

ndicators in Table 3A do not concern the same RPs, as best cases are selected separately in the HSN and LINKS. Identical RPs are selected for both the HSN and LINKS in *Table 3B*. All indicators with a * are categorical. Hence, N (mean) represents the number and the proportion of RPs belonging to that sub-group. Marriages are based on the population registers in the HSN. Using marriage certificates provides the following numbers in *Table 3A*: a. N= 194 mean=87.8, b. N=27 mean=12.2, c. N=137 mean=28.4, d. N=30 mean=13.3. Reading example: row 7 shows that 122 persons lived beyond the age of 30 years and married at some point in their lives in the HSN and that this was the case for 124 in LINKS. These respectively reflect 55.2% of the 221 persons who ever married in the HSN and 84.9% of the 146 persons who ever married in LINKS. Using the HSN marriage registers results in the following numbers is *Table 3B*: a. N=120 mean=87.0, b. N=18 mean=13.0, c. N=177 mean=27.7, c* N=177 mean=26.3, d. N=23 mean=13.0, d* N=26 mean=14.7. Reading example: row 7 shows that 138 persons lived beyond the age of 30 years in both the HSN and LINKS. Out of those 138 persons, 62 persons married at some point in their lives in the HSN and 118 in LINKS. These respectively reflect 44.9% and 85.5% of the in total 138 persons identified in both the HSN and LINKS. For siblings and children the N refers to the number of RPs with identified siblings. The mean refers to the mean number of siblings or children.

Comparisons of demographic indicators in the HSN and LINKS

Table 3A shows that the mean number of siblings and birth order were lower in the HSN (3.9 and 1.8) than in LINKS (6.7 and 4.2). These results were similar for the 186 identical cases. These differences are mainly a consequence of the research design of the HSN, in which siblings are only observed if they are part of the RPs' household. Therefore, information on siblings who died young or who did not live in the household is often missing, leading to an underestimation of sibship size in the parental household of the RP.

Within LINKS, marriages were available for 84.9% of the RPs of 30 years and older, whereas in the population register release of the HSN, marriages were available for 55.2% of the selected RPs. *Table 3B* shows that for the 138 common RPs, marital information was available for 85.5% in LINKS and for 44.9% in the HSN, which indicates that marriages were often not included in the population registers. Without selections, the mean age at marriage in the civil certificates and the population registers were 26.3 and 28.4, respectively. The higher mean age at marriage in the HSN is partly caused by right-censored observations in LINKS. Out-migration is known to cause underestimation of the number of events as well as the age at which demographic events occur[15]. Nevertheless, the number of known marriages is higher in LINKS than the HSN after we selected only individuals who married in Zeeland. The mean age at first marriage in LINKS was, at 26.8 years, higher than in the LINKS only selection as shown in *Table 3A*. In the HSN the age at first marriage in *Table 3B* was lower than in *Table 3A*, at 27.7 years. The higher age at first marriage in the HSN *(Table 3B)*, may be related to left truncation in the HSN, as not all RPs were followed for their entire life course, so that second marriages were counted as a first marriage, resulting in an overestimation of the mean age at first marriage. After combining the HSN population registers with the HSN marriage certificates, we observed that marriages were available for 87.0% of the RPs with a mean age at first marriage of 27.7 years (see the notes to *Table 3*). There is no evidence that passive registration lead to biased estimates. Differences between both datasets originate from registration procedures and censoring due to migration.

Table 3A shows that the mean number of identified children in the families of the RPs was similar between both datasets: 5.4 children for RPs in the HSN and 5.2 in LINKS. However, the number of RPs with identified children was lower in LINKS (N=152) than in the HSN (N=196). Furthermore, the mean ages at first and last birth in LINKS (26.5 and 36.6 years) were lower than in the HSN (27.0 and 37.4 years). The percentage of married couples without identified children is 14.6% in LINKS and 9.5% in the HSN. These differences are probably caused by right-censored observations in LINKS due to out-migration. *Table 3B* shows that for the 146 RPs who are included in both datasets, the mean age at first childbirth was 26.6 in LINKS and 26.8 in the HSN. This selection of common cases also showed the same mean number of children (5.4), although the mean age at last birth is lower in LINKS than in the HSN and the percentage of married couples without identified children is 15.5% in LINKS and 12.9% in the HSN. Apparently, the automatic linking procedure failed up to pick up specific certificates. Later-born children and entire families might be missing, as differences in mean age at last birth and mean number of children remain after selecting identical RPs.

The HSN and LINKS include different information on migration behaviour, as out-migration from Zeeland was not observed in LINKS. *Table 3A* shows that, according to the HSN, 95 (40.3%) of the RPs who were alive at age 18 migrated out of the province at some point in their lives. In LINKS, 157 death certificates are available for the RPs who lived at least until age

18, suggesting that these RPs either never left Zeeland or returned to Zeeland at a later age. The HSN indicated that 140 RPs (59.6%) never lived outside Zeeland, of which 62 (26.3%) never moved at all, and 78 (33.1%) only moved within Zeeland. Vital events outside the place of birth of the RP, indicating migration between municipalities within Zeeland, were identified for 67 RPs. This pattern was similar when we compared identical individuals *(Table 3B)*. In LINKS, we observed that 63 RPs (42.3%) who were observed after age 18, died in another municipality than their municipality of birth, whereas both vital events occurred in the municipality of birth for 86 RPs (57.7%). According to the HSN, 56 RPs (37.6%) remained in their municipality of birth, 71 (47.7%) moved within Zeeland and 21 (14.1%) lived outside Zeeland at some point in their lives. As about one out of seven adults who were born and died in Zeeland lived outside Zeeland at some time, assumptions about cross-provincial migration behaviour or the lack thereof should not be based on the presence of a death certificate in LINKS alone. The passive registration of individuals in the source material of LINKS means that migration movements can easily be missed.

The last rows in *Table 3A* present the number and mean age of death for all RPs for whom an age at death was known and the mean ages at death for individuals reaching 18 and 50 years. Because persons out-migrating from Zeeland are known in the HSN but not in LINKS, we expected that in the HSN more ages at death would be known and that the mean age at death would be higher. Indeed, fewer ages at death were known for RPs in LINKS than in the HSN, resulting in a lower mean age at death in LINKS (34.7 years) than in the HSN (40.8 years). The difference between the databases in the mean ages at death was smaller at higher ages. For those surviving until age 18, the mean age at death was 67.4 in LINKS and 69.4 in the HSN; after age 50, the mean ages at death are 73.6 and 75.1. An important reason for the declining difference with age is the declining likelihood with age that individuals will out-migrate (Kok, 1997). Differences between the HSN and LINKS are mitigated after identical cases were selected, which supports our assumption that selective availability of information for out-migrated individuals plays an important role. Hence, passive registration itself does not seem to cause biases in mortality estimates.

Comparing life-course and family reconstructions of RPs between the HSN and LINKS

Here, we take a closer look at deviations in individual life-course and family reconstructions. *Figure 2* shows whether estimations of outcomes in the HSN and LINKS deviate upward, downward, or are identical. Because information may be more complete for some subsets of individuals, four groups are included: individuals with a) no selections, b) known marriage certificates for parents of siblings and known death certificates for RPs, c) known migration inside Zeeland, and d) known migration outside Zeeland. Different estimations are seen as indicative of missed observations, as the matching procedure in both datasets has a low chance of producing false positives.

The HSN misses siblings that were not living in the RP's household. Without any selections on the data (Panel 2A), the number of siblings was higher for 69% of the RPs in LINKS, whereas 16% of the RPs in the HSN contained more siblings. However, LINKS also contained missed observations. The differences between the number of siblings in the HSN and LINKS were even more pronounced if a marriage certificate of the parents was known in LINKS (Panel 2B). Out-migration partly explains why family reconstructions in LINKS are better when a marriage certificate is available.

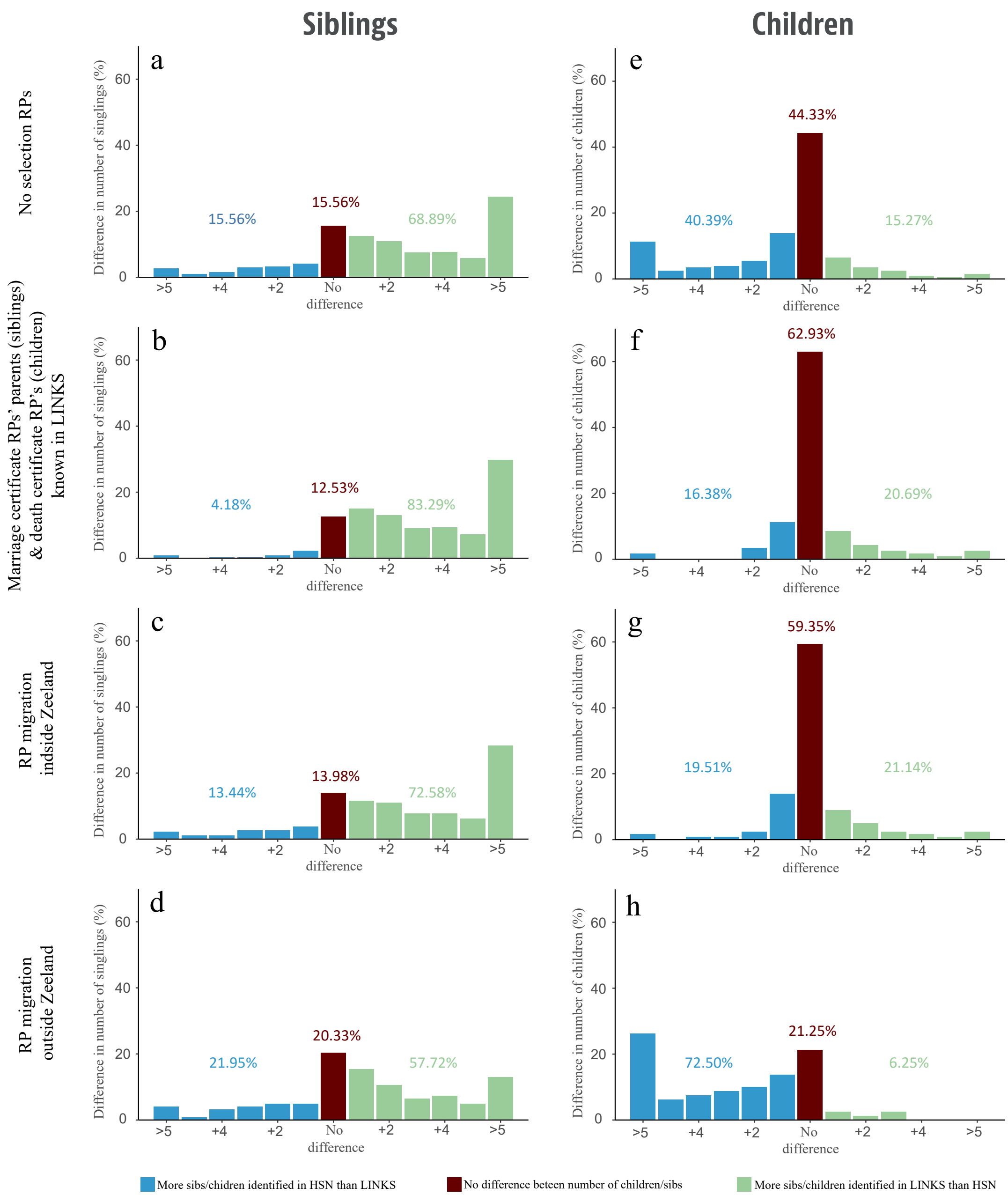


Figure 2: Matching for number of siblings and children between the HSN and LINKS.

The figure shows the matching for siblings and children of research persons between HSN and LINKS. Colors: red/black=exact match, blue/dark grey=more siblings or children identified in HSN than in LINKS, green/light grey=more siblings or children identified in LINKS than in HSN. Color descriptions first state the color version color and subsequently state the black and white version color (color/black and white). The x-axis show how much more siblings or children have been identified in either LINKS (green) or HSN (blue). The y-axis shows the percentage of matches corresponding to the x-axis. The marginal sums are illustrated with colors corresponding to the bars and sum up to 1 (100%). The legends on the left illustrate different data selections based on HSN or LINKS if explicitly stated. Numbers (N) per panel: A=495, B=407, C= 372, D=123, E=203, F=116, G=123, H=80.

In LINKS, fewer siblings were found in 4% of all cases compared to the HSN, the same number of siblings was found in 13% of all cases, and more siblings were found in 83% of all cases. Migration within Zeeland did not affect the results (Panel 2C), whereas for RPs who migrated out of Zeeland, the number of siblings in LINKS was lower than in the HSN in 22% of all cases, identical in 20% of the cases, and higher in the remaining 58% (Panel 2D). The availability of a parental marriage certificate is an independent observation that hints at successful matches between parents and their children. In general, reconstructions of sibships can be considered complete if such an independent observation is available.

The number of children of the RPs was more similar between both datasets than the number of siblings. *Figures 2E-H* show the difference in number of children between the HSN and LINKS, which was calculated for RPs who had children identified in either or both datasets. The active registration in the source of the HSN initially returned better results than the passive registration in LINKS. Without selections on the data, the HSN provided the most accurate results. For 40% of all RPs more children were found in the HSN than in LINKS; for 44% of all RPs the same number of children was found in both datasets; and in the remaining 15%, more children were found in LINKS than in the HSN. The difference between family size in the HSN and LINKS may have been caused by interprovincial migration, as births outside Zeeland are not included in LINKS. To indicate the quality of the linking process, RPs were selected who were known to have married, had children and died in Zeeland. The availability of a Zeeland death certificate for the RP and at least one Zeeland certificate for his children indicates that the RP spent a large part of his or her life in the province, thus reducing the chances that the RP migrated to a minimum. For these RPs, the same number of children was found in the HSN and LINKS in 63% and 59% of all cases respectively (Panels 2F and 2G). Where the number differed between the HSN and LINKS, there was no clear distinction in performance between the databases: the HSN performed better in some cases, whereas LINKS performed better in the others. If RPs moved out of Zeeland, a larger number of children was found in the HSN in 73% of all RPs, the same result in both sets in 21% of all cases, and a smaller number in the other 6%. Hence, the differences in family size between the HSN and LINKS are caused by migration rather than linking quality. Thus, the availability of a death certificate in LINKS indicates that observations on childbirth are likely available as well. This shows that passive registration can approach the quality of active registration when a later observation is available, e.g. a death certificate.

Figure 3 shows comparisons in the available mortality information between both datasets in four panels. The HSN returns more observations than LINKS, whereas the quality of matches is highly similar between both datasets. Panel 3A shows that ages at death were known for 409 RPs (83%) in the HSN, whereas 313 RPs had an available age at death in LINKS (63%). The age at death overlapped in 304 cases (99%) for whom a death certificate was available in both databases. Panel 3B presents childhood mortality for the RPs, their siblings, and their children. In the HSN, childhood mortality for RPs was estimated to be 6-7% higher than in LINKS, reflecting the good coverage of RP information in the HSN. However, observations on sibling and offspring mortality are of lower quality. Childhood mortality was estimated to be twice as high for siblings and almost three times as high for children in LINKS compared to the HSN. Panel 3C shows that adult mortality estimates were influenced by migration outside Zeeland. Among individuals who stayed in their municipality of birth or who moved within Zeeland, the mean and median ages at death were similar between both datasets.

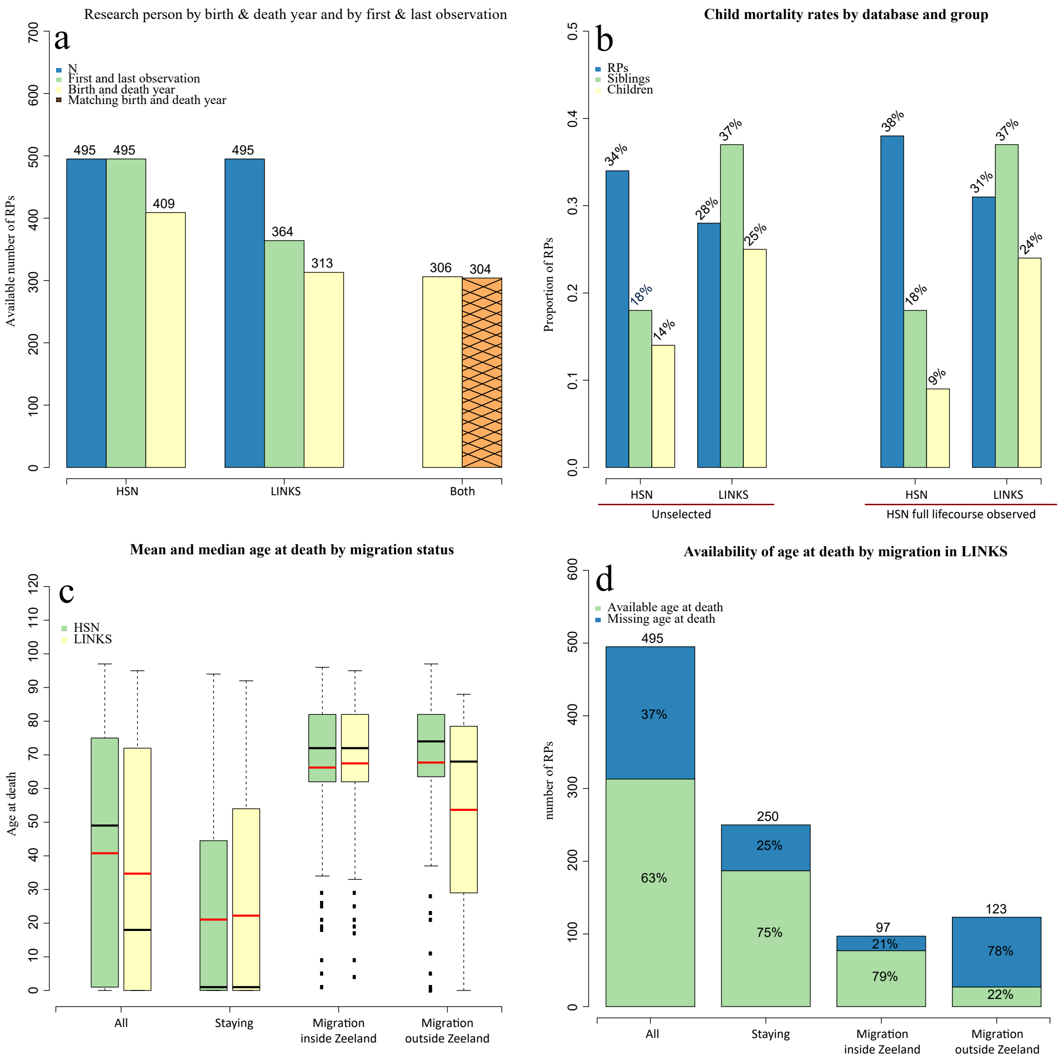


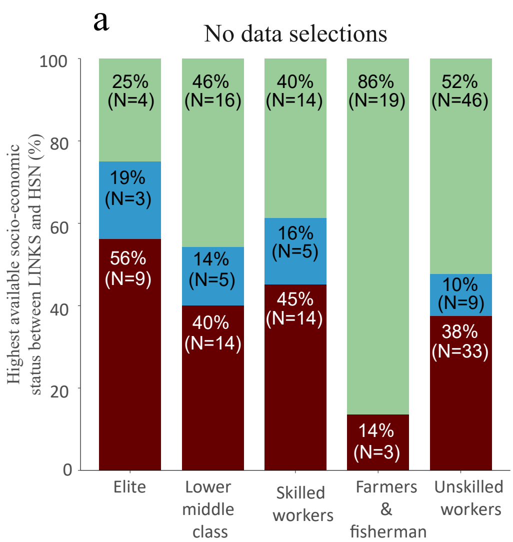
Figure 3: Comparison between HSN and LINKS for mortality data.

Panel A shows the research persons by birth and death year and by first and last observations (HSN only) in absolute numbers. Panel B shows the percentage of childhood mortality (mortality <5 years) by database (HSN and LINKS) and group (unselected and full life course). Full life course indicates that HSN RP's are observed from birth. Panel C shows the mean (red/light grey) and median (black) age at death by migration (staying, migration inside Zeeland, and migration outside Zeeland) by database (HSN and LINKS). Panel D shows the availability of age at death by migration only for the LINKS database in absolute numbers and percentages. Migration for RP's is determined based on the HSN since migration in LINKS is not available by definition of the source material.

However, death certificates were not linked for 25% of the individuals who were marked as stayers in the HSN and 21% of the individuals who were identified as internal migrants. Some of these individuals might have survived the observation period, as death certificates are not available after 1962. For other cases the death certificates were not linked due to spelling and age variations on the death certificates. There also is evidence of a "salmon bias" effect. For RPs who left Zeeland according to the HSN, the mean and median age at death is lower in LINKS than in the HSN. The date of death is known for only 22% of all RPs for whom we know, based on the HSN, that they lived outside Zeeland at some point in their life course (Panel 3D). These return migrants have a significantly lower age at death than interprovincial migrants who died outside Zeeland. Thus, passive registration returns fewer observations, but we find no proof for systematic biases related to the linking process. In addition, LINKS contains a selective group of stayers and return migrants, of which especially the latter may affect population estimates.

Figure 4 shows the differences in socioeconomic position between the HSN and LINKS based on the HISCLASS[35]. We present social class on the abbreviated HISCLASS scale with 5 categories: 1. elite, 2. lower middle class, 3. skilled workers, 4. farmers and fishermen, and 5. unskilled workers[10]. *Figure 4* further shows whether RPs in the HSN with an available HISCLASS-5 score had none, the same or a different score in the LINKS dataset. In general, the active registration in the HSN returned more cases than passive registration in LINKS. Panel 4A shows that in total 73 RPs - 33 women and 40 men - had known socioeconomic information in the HSN but not in LINKS. Conversely, 32 RPs - 25 women and 7 men - were recorded in LINKS, but not in the HSN. The share of missing values varied between 38% and 45% for unskilled workers, skilled workers, and the lower middle class, was slightly higher for the elite with 56%, and only occurred for 14% of the farmers. *Figure 4B* shows that HISCLASS scores were identical for 80% of the RPs for whom occupational information was known in both datasets. All farmers in the HSN were also identified as farmers in LINKS. However, differences in social position were found for 22% of the other RPs. Most discrepancies with the HSN occurred for the elite (43%), more than for the lower middle class (24%) and skilled workers (29%). Fewer differences with the HSN were found for the unskilled workers (16%). Underestimation of socioeconomic status generally occurred when information on occupational status was not known after marriage (Delger & Kok, 1998). These problems with censoring are probably caused by migration, rather than passive registration in the source. Geographic mobility is known to be higher for individuals with a better socioeconomic position[39], so that observations of those who reach a higher social position in society are more likely to be censored. Therefore, local datasets underestimate the social position of migrants as less occupational information is available at higher ages, and are biased towards stayers who, on average, reach a lower social standing.

Figures 4C-E show comparisons of the occupational score in the HSN with the LINKS score on the RP's death certificate, his or her marriage certificate, and the marriage and death certificates of the RP´s children. The choice for a certain certificate determined the sample size. Occupations were only recorded on death certificates if the deceased held an occupation at the time of death. As a result, occupational information on death certificates is limited and only available for 29 cases, but the HISCLASS scores were very similar between the datasets. Marriage certificates were available for 112 RPs, of whom 52 were identified as unskilled workers in the HSN. In 98% of the cases, these were also identified as unskilled labourers on their marriage certificates. However, marriage certificates are less concordant with the HSN for socially mobile individuals. Between 36% and 42% of the farmers, skilled labourers, and lower middle class had a different occupational position on their marriage certificate than in the HSN. This difference was larger for the elite (57%). The 59 RPs with marriage and death certificates of children in the LINKS dataset (Panel 4E) have a better balance between sample size and matching quality in socioeconomic position than the comparison made in Panel 4D. Similarly, farmers show no differences at all between the HSN and LINKS datasets. For the other groups, socioeconomic positions differ from 23% to 33% of unskilled workers, skilled workers, and the lower middle class. For the elite two out of the three observations are different. Individuals with more children have more observations of their socioeconomic status, and for them HSN and LINKS reflect each other better. More generally, because in passive registration databases the number of observations depends on the number of linked events, passive registration databases reflect the active registration database better when more events were linked.



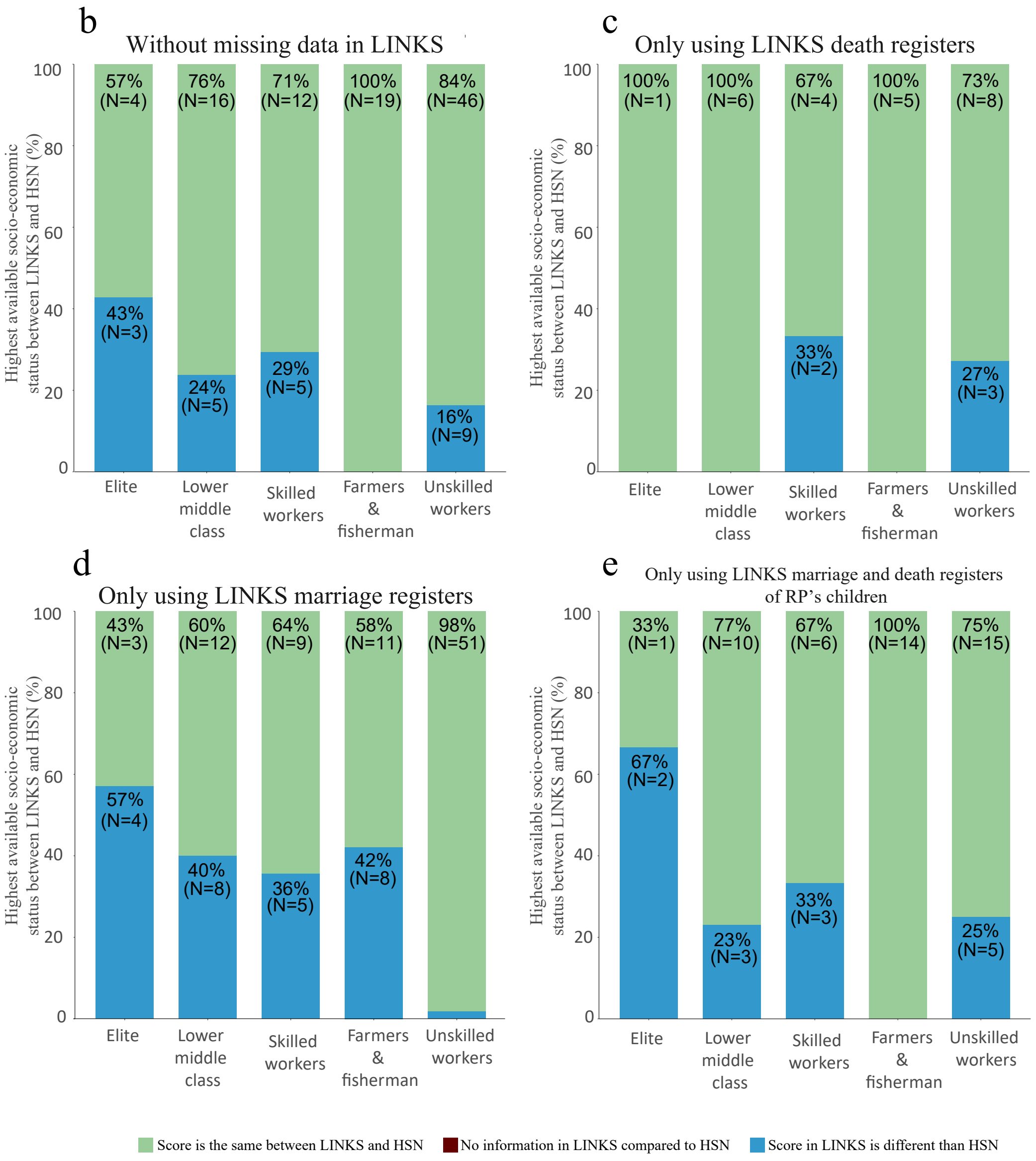


Figure 4: HISCLASS score for research persons in HSN and LINKS.

32 RP's are available in LINKS and not in HSN (7 males, 25 females. 119 RP's are available in both the HSN and LINKS. (91 males, 28 females). Panel A shows the proportion of highest available socio-economic status between LINKS and HSN without any data selection LINKS. Panel B shows the proportion of highest available socio-economic status between LINKS and HSN without the 73 missings in LINKS. Panel C shows the proportion of highest available socio-economic status between LINKS and HSN with only information of death certificates used in LINKS. Panel D shows the proportion of highest available socio-economic status between LINKS and HSN with only information of marriage certificates used in LINKS. Panel D shows the proportion of highest available socio-economic status between LINKS and HSN with only information of marriage and death certificates of RP's children used in LINKS.

Conclusion and discussion

In this paper, we compared life course and family reconstructions for 495 individuals who are available in two different types of data sources: the HSN based on active registration in the population registers, and LINKS, based on passive registration from the civil certificates. We found that differences between the HSN and LINKS were caused by censoring due to migration, rather than the nature of the administrative process which seems to induce more random missings. The established practice of selecting specific cases[1] made most differences in demographic estimates between the databases based on active and passive registration disappear, but only for demographic estimates at the individual level.

In general, the identification of children appears to be more complete when databases are based on active registration. The total number of families with children as well as the number of identified children per family is higher in the HSN than in LINKS. However, after adjustments to exclude inter-provincial migration, the number of children identified in the databases was usually identical between the databases. This finding illustrates that the identification of children using passive registration is of similar quality to active registration for non-migrants. In line with our expectations, the number of RPs with known siblings or the size of the RPs' sibling set is smaller in the HSN than in LINKS. Sibling reconstructions in LINKS are complete when a marriage certificate of the parents is available. Due to the research design of population registers and the HSN, not all siblings were found in the population registers in which the RPs appeared. Apart from missed migrants, LINKS seems to contain well-reconstructed families, so that not only the number of children, but also the siblings are identified in the dataset. For both databases based on sources with passive and active registration, it seems best to only include observations on siblings or offspring when separate observations indicate that observations are not censored.

Population estimates on demographic behaviour are strongly affected by whether observations are missed due to migration. Ruggles (1992) used simulation methods to show that – even in the absence of healthy migrant effects – cessation of observation on individuals due to out-migration causes underestimation of the ages at which demographic events occur[15]. Later work pointed out that migration at young ages, or because of a marriage at the same age as in the population of origin, do not bias estimations of age at marriage[40]. As more individuals were lost from observation due to migration in LINKS than in the HSN due to the provincial scope of LINKS and the national scope of the HSN, we expected that the mean age at which life-course transitions occur would be lower in LINKS than in the HSN. Indeed, we found that not only age at death, but also age at first marriage, first childbirth, and last childbirth were higher in the HSN than in LINKS. More generally, this implies that mean estimates, such as average age at death of a study population, show a stronger downward bias when the loss of observation due to migration increases. However, estimates of age at marriage are much less affected by migration. This contrasts earlier work which showed that in some populations migration patterns may not distort estimations of age at marriage altogether, as individuals migrating out may migrate for marriage specifically, or very early in life, before they are at risk of marrying[14,40]. In addition, in LINKS, more men and women had no identified children than in the HSN. Censoring of observations due to migration - and not passive registration - thus has a significant effect on population estimates.

Ages at death were identical in the HSN and LINKS for 304 out of 306 cases, indicating the validity and comparability of the life-course reconstructions in both databases. In line with earlier observations from Hacker (1997), migration seemed to have a strong effect on mortality estimates[13]. We expected that we would find a lower mean age at death in LINKS than in the HSN, as it has been shown that migrants are often healthier than the native population, a phenomenon known as the healthy migrant effect. Indeed, we found a lower mean lifespan for the RPs in LINKS than in the HSN, which was attributable to the almost 100 extra observations of lifespans that were available in the HSN in comparison to LINKS. These observations mainly concerned out-migrated adults, increasing the length of the mean life span in the HSN. Moreover, we found that individuals who were observed outside the province of Zeeland during their life course, but who returned to Zeeland, died at earlier ages than individuals who never migrated or who migrated within the province of Zeeland. This suggests that return migration occurred for health considerations, contributing to the problem of underestimation of ages at death in LINKS. In sum, this means that reliable estimates on mortality rates in the general population cannot be derived from regions with pronounced out-migration, unless subgroups are studied (e.g. infants or those 50+) or moments of censoring after the last observation are inferred[4,5,16,41]. However, one can wonder how useful the latter method is, seen that it only corrects mortality estimates for when individuals migrate, i.e. 15-50, and not for when migrants have left, ages 50 and up.

3

In the literature, a number of earlier studies have reported findings in line with the "salmon bias", which states that the relative health advantage of migrants in comparison to the native population may at least partially be caused by return-migration movements of unhealthy migrants. Earlier work has found that healthier individuals tend to migrate more and further in contemporary as well as historical populations[42,43]. Work from England has shown that migrants affected by pulmonary tuberculosis tended to return to their regions of origin, leading to high mortality rates in sending regions and relatively low mortality rates in receiving regions[44]. At the same time, a historical study on Rotterdam did not find evidence for either a healthy migrant effect or a salmon bias[45] Evidence for the current data is in line with both healthy migrant effects and a salmon bias. Possibly, in Rotterdam, healthy migrant effects were counterbalanced by a heavy urban penalty affecting migrant's health, which is absent in the small towns of Zeeland. Alternatively, for salmon bias to occur a disease has to be in chronic rather than a short sickbed before death[46]. The occurrence of salmon bias may therefore be related to spatial differences in disease patterns.

This paper has illustrated that life-course and family reconstructions based on linked, passive registration on individuals constitute a reliable alternative to such reconstructions based on active registration. Through the further integration of existing sources databases for innovative new research may be generated. Information from different datasets can be combined to gain new and more complete insights into demographic behaviour. The extensive family networks found in LINKS can contribute to more detailed kinship information in the HSN; for instance, with regard to lifeless reported infant siblings and children or more detailed observations on socioeconomic status. In current versions of the HSN, marriage certificates - which are also included in LINKS - are already used to enrich information on relationship formation found in population registers. Second, differences between the two databases may itself be of interest for family historians and historical demographers. Deviating information on siblings and children within households in the HSN and regardless of household in LINKS may provide researchers with clues on non-co-resident kin, a phenomenon on which neither database alone provides information. Similarly, supplemental observations on socioeconomic status in the HSN may enrich our understanding of the development of the status of individuals over time. As the current analyses have shown, it should be taken into account that information for certain individuals may more readily be matched between databases, such as index persons from LINKS who remained in their province of origin.

References

1 Bousquet, G.-H. & Henry, L. Manuel de démographie historique. *Rev. économique* **21**, 513 (1970).

2 RUGGLES, S. The limitations of English family reconstitution: English population history from family reconstitution 1580–1837. *Contin. Chang.* **14**, 105–130 (1999).

3 Song, X. & Campbell, C. D. Genealogical Microdata and Their Significance for Social Science. *Annu. Rev. Sociol.* **43**, 75–99 (2017).

4 Gill, R. D. Nonparametric estimation under censoring and passive registration. *Stat. Neerl.* (1997). doi:10.1111/1467-9574.00036

5 Alter, G., Devos, I. & Kvetko, A. Completing Life Histories with Imputed Exit Dates: A Method for Historical Data from Passive Registration Systems. *Popul. (english Ed.* (2009). doi:10.3917/pope.902.0293

6 Wisselgren, M. J., Edvinsson, S., Berggren, M. & Larsson, M. Testing methods of record linkage on Swedish censuses. *Hist. Methods* (2014). doi:10.1080/01615440.2014.913967

7 Delger, H. & Kok, J. Bridegrooms and biases: A critical look at the study of intergenerational mobility on the basis of marriage certificates. *Hist. Methods* (1998). doi:10.1080/01615449809601194

8 Gavrilov, L. A. & Gavrilova, N. S. Biodemographic Study of Familial Determinants of Human Longevity. *Popul. An English Sel.* **13**, 197–221 (2001).

9 Mandemakers, K. Building life courses datasets from population registers by the historical sample of the Netherlands. *in History and Computing* **14**, 87–107 (2002).

10 Kok, J., Mandemakers, K. & Bras, H. Van geboortebank tot collaboratory. Een reflectie op twintig jaar dataverzameling en onderzoek met de HSN. *Tijdschr. voor Soc. en Econ. Geschiedenis/ Low Ctries. J. Soc. Econ. Hist.* **6**, 3 (2009).

11 Schellekens, J. & van Poppel, F. Early-life conditions and adult mortality decline in Dutch cohorts born 1812–1921. *Popul. Stud. (NY).* (2016). doi:10.1080/00324728.2016.1223336

12 Adams, J. W., Kasakoff, A. & Kok, J. Migration over the life course in XIX th century netherlands and the American north: A comparative analysis based on genealogies and population registers. *Ann. Demogr. Hist. (Paris).* (2002). doi:10.3917/adh.104.0005

3

13 Hacker, J. D. Trends and Determinants of Adult Mortality in Early New England: Reconciling Old and New Evidence from the Long Eighteenth Century. *Soc. Sci. Hist.* (1997). doi:10.1017/s014555320001782x

14 Voland, E. & Dunbar, R. I. M. The impact of social status and migration on female age at marriage in an historical population in north-west Germany. *J. Biosoc. Sci.* **29**, 355–360 (1997).

15 Ruggles, S. Migration, Marriage, and Mortality: Correcting Sources of Bias in English Family Reconstitutions. *Popul. Stud. (NY).* **46**, 507–522 (1992).

16 Jonker, M. A. & van der Vaart, A. W. Correcting missing-data bias in historical demography. *Popul. Stud. (NY).* **61**, 99–113 (2007).

17 Kasakoff, A. B. & Adams, J. W. The effect of migration on ages at vital events: A critique of family reconstitution in historical demography. *Eur. J. Popul.* (1995). doi:10.1007/BF01264948

18 Thorvaldsen, G. Using NAPP Census Data to Construct the Historical Population Register for Norway. *Hist. Methods A J. Quant. Interdiscip. Hist.* **44**, 37–47 (2011).

19 van Dijk, I. K. *Death and the family. High mortality families and the life course, The Netherlands, 1812-1912.* (Enschede: Ipskamp, 2019).

20 Massey, C. G. Playing with matches: An assessment of accuracy in linked historical data. *Hist. Methods A J. Quant. Interdiscip. Hist.* **50**, 129–143 (2017).

21 Ruggles, S., Fitch, C. A. & Roberts, E. Historical Census Record Linkage. *Annu. Rev. Sociol.* (2018). doi:10.1146/annurev-soc-073117-041447

22 Vulsma, R. F. Burgerlijke stand en bevolkingsregister *[Civil Register and Population Register].* (Centraal Bureau voor Genealogie, 1988).

23 van den Berg, N. et al. Longevity defined as top 10% survivors and beyond is transmitted as a quantitative genetic trait. *Nat. Commun.* **10**, 35 (2019).

24 Priester, P. *Geschiedenis van de Zeeuwse landbouw circa 1600-1910.* ('t Goy-Houten: HES Uitgevers, 1998).

25 Knotter, A. & Meijer, A. C. De gemeentelijke bevolkingsregisters, 1850-1920. *Den Haag Inst. voor Ned. Geschied.* (1995).

26 Stadsarchief Amsterdam (2018). Handleiding Gezinskaarten 1893-1939. Available at: https://archief.amsterdam/indexen/gezinskaarten_1893-1939/handleiding/index.nl.html.

27 Mandemakers, K. Historical sample of the Netherlands. In P. K. Hall, R. McCaa, & G. Thorvaldsen (Eds.). in *Handbook of International Historical Microdata for Population Research* 149–177 (2000).

28 IISG. (2018). About HSN: Sampling. Retrieved from https://socialhistory.org/en/hsn/sampling.

29 Janssens, *A. Family and social change. Canadian Journal of Sociology / Cahiers canadiens de sociologie* (Cambridge University Press, 1993). doi:10.1017/CBO9780511628696

30 Mandemakers, K. & Laan, F. LINKS dataset Genes Germs and Resources. WieWasWie Zeeland. Civil Certificates. (2017).

31 *HSN (2010b). Historical Sample of the Netherlands (HSN): Data set life courses release 2010.01 [Data file and code book].*

32 *HSN (2010a). Historical Sample of the Netherlands (HSN): Data set civil certificate release 2010.01 [Data file and code book].*

33 Mandemakers, K. https://socialhistory.org/en/hsn/hsn-releases. *HSN 2010.01 release* (2010).

34 Fu, Z., Boot, H. M., Christen, P. & Zhou, J. Automatic Record Linkage of Individuals and Households in Historical Census Data. *Int. J. Humanit. Arts Comput.* (2014). doi:10.3366/ijhac.2014.0130

35 Van Leeuwen, M. H. D. & Maas, I. *HISCLASS: A historical international social class scheme.* (Universitaire Pers Leuven, 2011).

36 Mandemakers, K. et al. HSN standardized HISCO-coded and classified occupational titles. (2018).

37 Janssen, F. & van Poppel, F. The Adoption of Smoking and Its Effect on the Mortality Gender Gap in Netherlands: A Historical Perspective. *Biomed Res. Int.* **2015**, 1–12 (2015).

38 van dijk, I. K. & Mandemakers, K. Like mother, like daughter. Intergenerational transmission of infant mortality clustering in Zeeland, the Netherlands, 1833-1912. *Hist. Life Course Stud.* 1–26 (2018).

39 Ekamper, P., van Poppel, F. & Mandemakers, K. Widening Horizons? The Geography of the Marriage Market in Nineteenth and Early-Twentieth Century Netherlands. in *Navigating Time and Space in Population Studies* (2011). doi:10.1007/978-94-007-0068-0_6

40 Desjardins, B. Bias in Age at Marriage in Family Reconstitutions: Evidence from French-Canadian Data. *Popul. Stud. (NY).* **49**, 165–169 (1995).

41 Jonker, M. A. & Van Der Vaart, A. W. A semi-parametric model for censored and passively registered data. *Bernoulli* (2001). doi:10.2307/3318600

42 Alter, G. & Oris, M. Childhood conditions, migration, and mortality: migrants and natives in 19th-century cities. *Soc. Biol.* (2005). doi:10.1080/19485565.2005.9989108

43 Lassetter, J. H. & Callister, L. C. The Impact of Migration on the Health of Voluntary Migrants in Western Societies: A Review of the Literature. *J. Transcult. Nurs.* (2009). doi:10.1177/1043659608325841

44 Hinde, A. Sex differentials in phthisis mortality in England and Wales, 1861–1870. *Hist. Fam.* (2015). doi:10.1080/1081602X.2015.1051077

45 Puschmann, P., Donrovich, R. & Matthijs, K. Salmon Bias or Red Herring? *Hum. Nat.* **28**, 481–499 (2017).

46 Reid, A. & Garrett, E. Mortality, Work and Migration. A Consideration of Age-specific Mortality from Tuberculosis in Scotland, 1861-1901. *Hist. life course Stud.* **6**, 111–132 (2018).

47 Campbell, C. D. Demographic Techniques: Family Reconstitution. *in International Encyclopedia of the Social & Behavioral Sciences* 138–142 (Elsevier, 2015). doi:10.1016/B978-0-08-097086-8.31012-1