Kees Mandemakers

Life Course research with the Historical Sample of the Netherlands (HSN). [1]

Working paper HSN 42 (IISG 2017).

1. Why are Life Course Data useful?
What kind of people live in their place of birth during their whole life? And who move away to a bigger city, another area or go all the way and migrate to another country? And do these persons ever come back or do they stay away for good? Under which circumstances do children grow up and how high is the chance of a breakup of their parents' marriage and what would be the influence of a divorce on their chances of success in society? And what is the relationship between societal success and school certificates?

Everyone has asked himself these kinds of questions and they have been the subject of scientific studies: research with an important historical dimension. After all, societies change: a century ago marriages lasted unil one of the partners died, whereas nowadays most marriages end up in a divorce.[2] Which circumstances increased the risk of a premature loss of one's partner and did people remarry relatively often or seldom? And who was this new marriage partner? Was it another widow or widower, or was it a living-in servant or indispensable maid? Labour markets change: in the nineteenth century almost half of the Dutch population worked in agriculture and nowadays this is hardly two percent. Migration has been another factor for big changes in society. This is shown very clearly in Rotterdam, a city with more than a hundred nationalities. The urban environment has always been influenced by big changes. Not so long ago a high influx of Germans and people from Brabant and Zeeland into Rotterdam played a major role in the flourishing of the city of Rotterdam.[3] How rapidly did they integrate into society?

It is clear that a lot of present day questions concerning society played a role in the past. Of course, long-term research can only be done by using longitudinal data. These are data which are acquired when persons are followed for a long period. This type of investigation is done within the social sciences, epidemiology and history and is called 'life course research'.[4] Tracking life courses systematically and at a large scale shows long term processes and enables

---

[1] This paper is based on revised versions of Kees Mandemakers, *Waarom Jan en Cor met elkaar trouwden: Over grote historische databestanden, koudwatervrees en interdisciplinaire samenwerking* (Aksant: Amsterdam 2009); Kees Mandemakers, 'Levensloeponderzoek in Rotterdam met de Historische Steekproef Nederlandse bevolking (HSN)', in: Paul van de Laar, Leo Lucassen & Kees Mandemakers (red.), *Naar Rotterdam. Immigratie en levensloop in Rotterdam vanaf het einde van de negentiende eeuw.* (Aksant Amsterdam 2006); 9-24 and Jan Kok, Kees Mandemakers en Hilde Bras, 'Van geboortebank tot *collaboratory*. Een reflectie op twintig jaar dataverzameling en onderzoek met de HSN', *Tijdschrift voor Sociale en Economische Geschiedenis* 6 (2009), 4, 3-36.

[2] F. Van Poppel, *Trouwen in Nederland. Een historisch-demografische studie van de 19ᵉ en vroeg-20ᵉ eeuw* (Wageningen 1992), 278. *A.A.G. Bijdragen* 33.

[3] See a.o. P. van de Laar, L. Lucassen en K. Mandemakers (red.), *Naar Rotterdam. Immigratie en levensloop in Rotterdam vanaf het einde van de negentiende eeuw* (Aksant: Amsterdam 2006).

[4] M. Kalmijn, 'Sociologische analyses van levensloopeffecten: een overzicht van economische, sociale en culturele gevolgen', *Bevolking en Gezin* 31 (2002), 3, 3-46; J. Kok, 'Transities en trajecten. De levensloopbenadering in de sociale geschiedenis', *Tijdschrift voor Sociale Geschiedenis* 26 (2000), 309-329; A.C. Liefbroer en P.A. Dykstra, *Levenslopen in verandering. Een studie naar ontwikkelingen in de levenslopen van Nederlanders geboren tussen 1900 en 1970* (SDU Uitgevers: Den Haag 2000).

researchers to analyse the social, economic and cultural factors which produce these changes in their mutual relationship. The data for this kind of analyses are stored in large-scale longitudinal microdatabases.

In Rotterdam for example, the Generation R Study started in 2001. It is a prospective cohort study, which tracks 10,000 children from fetal life until their twentieth year. The study focuses on physical, behavioral and cognitive development, diseases and health care in a multi-ethnic urban population. The project centers around the question why some children develop in an optimal way and others don't.[5]

In order to understand these questions about the genetic and environmental causes of growth, development and health for the past as well, longitudinal data are needed. Interviewing or following persons year by year is not possible any more. For historical research into migration for example it is not sufficient to use aggregated data like the total sum of annual departures and arrivals. Research must be done in the archives themselves, tracking migrating persons and following them during their lives.

Fortunately there are lots of data for the period after 1800. The Historical Sample of the Netherlands (HSN) contains individual life courses from the nineteenth and twentieth century. In section 2 the life course approach will be explained by presenting the sources most used for this kind of databases. In section 3 the historical development of the HSN itself will be outlined and in section 4 some examples will be given of research with HSN data.

2. Sources of the HSN

*Registry of Births, Deaths and Marriages[6]*
The research subjects for the HSN-database (about 85,000 persons) were selected by way of an a-select sample from the registers of birth for the period 1812-1922.[7] The HSN systematically collects data about these persons from the Civil Registry of the Netherlands: the data are found in the certificates of birth, marriage and death. These certificates are kept at public archives. The birth certificates form the basis of the life course researches: they provide quite a number of data. Illustration 1 shows the birth certificate of Jan Hendrik Bruning, born on 8 February 1869 in Rotterdam at 10 pm. The certificate shows for example that his father, Jan Christiaan Willem Bruning, was capable of putting his signature on the certificate. This means that he must have attended school for a few years. His mother, Catharina Elizabeth Moerman, did not have an occupation, according to the certificate. The certificate also gives the address and the occupation of the father. Besides this, it also gives the names, ages and professions of the witnesses. Here we find that the father had brought at least one colleague to co-sign the certificate as a witness. Birth certificates also always state whether the mother of the new-born was married or unmarried at the time of birth.

[5] http://www.generationr.nl/; 25th of april 2017.
[6] For general information about the Dutch civil registration, see R.F. Vulsma, *Burgerlijke stand en bevolkingsregister.* 2e revised print, Den Haag, 2002.
[7] Mandemakers, K. (2000), Historical Sample of the Netherlands. In: P. Kelly Hall, R. McCaa en G. Thorvaldsen (red.), *Handbook of International Historical Microdata for Population Research.* Minneapolis: Minnesota Population Center, pp. 149-178.

Illustration 1. Birth Certificate Rotterdam, 1869, certificate number 544 [Jan Hendrik Bruning born on the 8th of February 1869, certificate date: 02-10-1869]
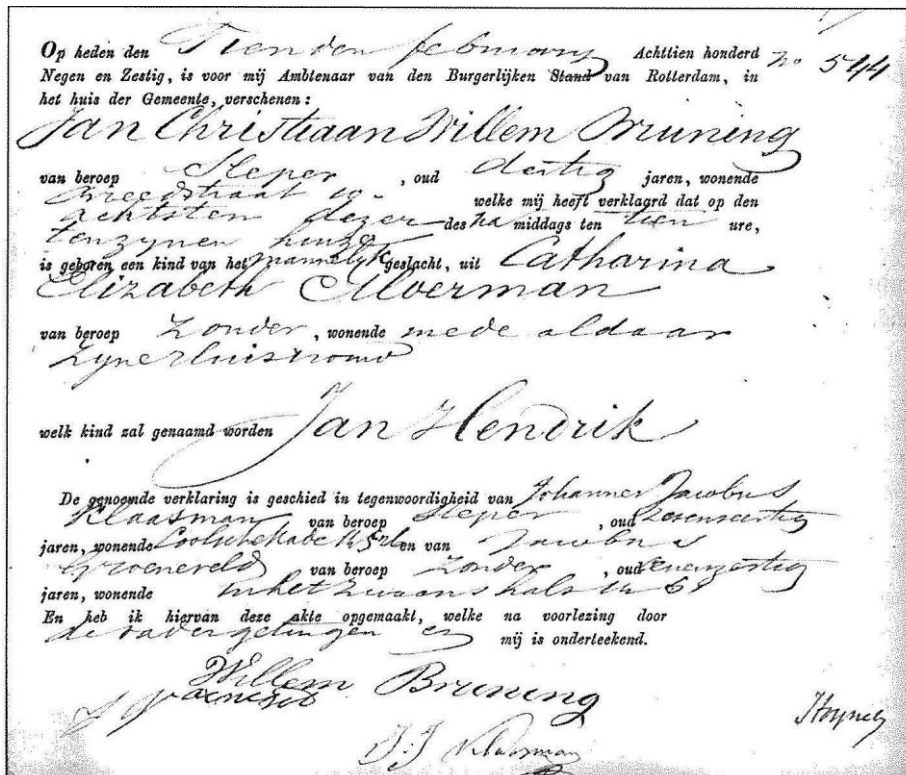
After the entry of the birth certificates into the HSN database, the certificates of death are searched for. Of course the most important data needed is the date of death, but these certificates also give some information about migration, whether the deceased person died in another municipality than where he or she was born. Usually it was the father who gave notice of the death of a child, which means that in the case of 'yound deaths' there is a second opportunity to find information about his occupation and whether or not he could write his signature, an indicator for analfabetism.

The next step for the HSN database is the entry of the data from the certificates of marriage. Illustration 2 shows the first marriage of the before mentioned Jan Hendrik Bruning. He had been fortunate enough to survive the risky years of early childhood and at the age of 24 he married Cornelia Stuip in 1893. She was a 20-year-old maid and came from Nieuw-Beijerland. The bridegroom was still living in Rotterdam and was a typesetter by profession. His father had died by that time and his mother was able to write her signature, which means she will have had some form of schooling. A second certificate of marriage – not included here- from 26 March 1924 shows that Jan Hendrik Bruning had become a widower and had remarried 41-year-old Jentje Dekker, a shopkeeper.

Marriage certificates are a very rich source. They do not only show the data of the marriage partners, but also data concerning their parents and their four witnesses. Usually the witnesses are relatives or friends of the man and woman who are getting married. But not in this case: all four of them are townsergeant, which is another term for a town messenger. Apparently they acted as a kind of professional witnesses here. The certificates provide data about several interrelated persons, concerning their dwelling places, ages, occupations,

signatures and relationships. In combination with other certificates a lot of research can be done with these kind of data.

Illustration 2. Marriage Certificate Rotterdam, 1893, certificate number 424 [Jan Hendrik Bruning married on the 26th of March 1893]

Population register[8]

The ultimate goal of the HSN is the reconstruction of the complete life course of the sampled persons. In order to do so all information that can be found in the population registers must be recorded. Illustration 3 shows an example of a family card. Here we meet Jan Hendrik Bruning again, with his wife and two children. His second child, Marinus Dirk, had died within two months. The card also tells us about his religious affiliation, Jan Hendrik was Dutch-Reformed, like the whole family. His occupation, typesetter, is affirmed here. This source also shows that Han Hendrik Bruning provided temporary housing to a brother-in-law, a sister-in-law, a brother and a cousin. Migration data are found in the columns place of origin and place of departure. We can tell that his son, Jan Christiaan Willem, left for Schoonhoven on 11 April 1910, to the military barracks situated there. And on 7 June 1910 his quick return is registered. On 3 October 1910 he left the household permanently and a family card bearing his name was made. The brother-in-law, Cornelis Stuip, a breadbaker, arrived on 25 January1898 from Nieuw-Beijerland and left for Harderwijk on 26 July 1898, he went to the barracks as well.

By linking all population registers of the more than a thousand municipalities of that time, it is possible to track down the complete migration pattern of all these persons. The composition of the family can be established when looking at the relation of the family members to the head of the family, e.g. 'son', 'cousin' or 'lodger'. On the basis of the migration data and the dates of birth and death it is possible to establish the composition of and the changes within the family. This family card shows for example that his first wife died on 17 July 1922. For his second marriage a new family card was made for Jan Hendrik, which is marked on the card with the term 'GK II' (gezinskaart=family card) in the column *Dag en Jaar van overlijden (day and year of death).*

The Netherlands is one of the few countries in the world with a continuous population register starting as early as the mid-19th century. In the early registers each household was entered on a double page, starting with the head of the household; he was followed by his wife, children, other relatives, and other members of the household. Date and place of birth, relation to the head of the household, sex, marital status, occupation, and religion were recorded for each individual. All changes occurring in the household were recorded in the register. Population registers remained in use until 1910 or 1920, after which a new form of continuous registration was introduced, consisting of single sheets, so-called family cards. From then on the registration unit was no longer the household, but the family.

The population register was introduced in 1850. In the course of time, in Rotterdam around 1890, larger municipalities changed from working with registers to working with family cards (Ill.3). When moving house the data concerning the family were not kopied onto another registerpage, instead there was only a change in the registration of the address. On this card we can find nine different addresses. Unfortunately it was not until 1920 that registers started to record the date of the change of address. On 1 October 1923 Jan Hendrik was the only one left in the household and he left for the Nieuwe Binnenweg 436b to live in with his brother-in-law.

---

[8] For a description of the population register, see Knotter, A en A. C. Meijer (red.) (1995). *De gemeentelijke bevolkingsregisters 1850-1920*. Den Haag: Instituut voor Nederlandse Geschiedenis and/or Schrover, M. en P. Vuijst, 'Familiereconstructies en bevolkingsregister', in: VULGO, *Hoe krijg ik mijn bronnen aan de praat?* (Utrecht 1990), 34-53.

Illustration 3. Family Card population register Rotterdam of Jan Hendrik Bruning born on the 8th of February 1869 and married to Cornelia Stuip.

It will be clear by now that not only the data of the sampled persons are reproduced from the registers, but also the data concerning the other persons in the same household. A typical life course is characterized by four different situations: a) growing up within the parental home, b) living in with other families or boarding, as a maid, schoolpupil, tradespupil, lodger, soldier, c) as a parent with a family of one's own and d) living in when one has become of age or in needy circumstances. Of course many persons did not go through all these phases, many of

them died as a child. And not every one got married or had any children, and some had already left the family home before getting married.

In the late 1930s, the population register was replaced by the personal card (PK); from that time on the individual person became the registration unit in all municipalities. When a person moved from one municipality to another, this PK moved with him. Since then the population register in each municipality has consisted of a collection of personal cards, containing nearly the same information as the population register. All persons who were alive in July 1938 or were born after that year received a personal card. At the time of death, this card is removed from the files and sent to the Central Bureau of Statistics (CBS), where the data on the card are used for statistical purposes; and finally the card is sent to the Central Genealogical Bureau (CBG). Copies of the cards have been used to enrich the data set of the HSN. They contain the following information: name, municipality and date of birth of the person concerned, as well as those of his or her parents, marriage partner(s) and children. The nationality is given as 'Dutch' or 'Foreign'. Successive occupations, addresses and changes therein are also indicated. The whole migration path (all adresses), the composition of the family and the religous affiliations can be found here.[9]

This system remained in use until 1 October 1994, when the card system changed into an electronic register, the so-called *Gemeentelijke Basis Registratie* (GBA). In the early 21st century the system was centralized on the national level and since 2014 it is known as the *Basis Registratie Personen* (BRP). As soon as a person has died a so-called Personal List with personal information is sent to to the Central Genealogical Bureau as a replacement for the Personal Card. This unique source enables research of social mobility, secularization, changes in the composition of the family and migration patterns.

3. The database of the HSN

The sample of the HSN is drawn from the birth certificates and stratified in periods of ten years. To achieve rather equally sized cohorts of persons from the age of twenty years, depending on infant and child mortality on the one hand and the number of births on the other hand, it was decided to have two sample frequencies: 0.75% for the period 1812-1872 and 0.5% for 1873-1922. This results in a sample size that is large enough to make sound statistical conclusions for subpopulations of minimal two percent of the persons born in the Netherlands during the 19th and early 20th century (in total about 14,5 million persons) at the age of 20.

Figure 1 gives an overview of the data gathered for each RP since the start of the HSN in 1991. In the first ten years the HSN concentrated on the data entry of all birth certificates and the death certificates of children who died before the age of ten. After the year 2000 more and more marriage certificates were entered and the HSN also started entering data from the population register.

---

[9] For an introduction to the personal cards, see C. Gordon, *The Bevolkingsregisters and their use in analyzing co-residential behaviour of the elderly* (Den Haag 1989), NIDI rapport 9.

Figure 1   Development of the HSN-database, 1991-2015.
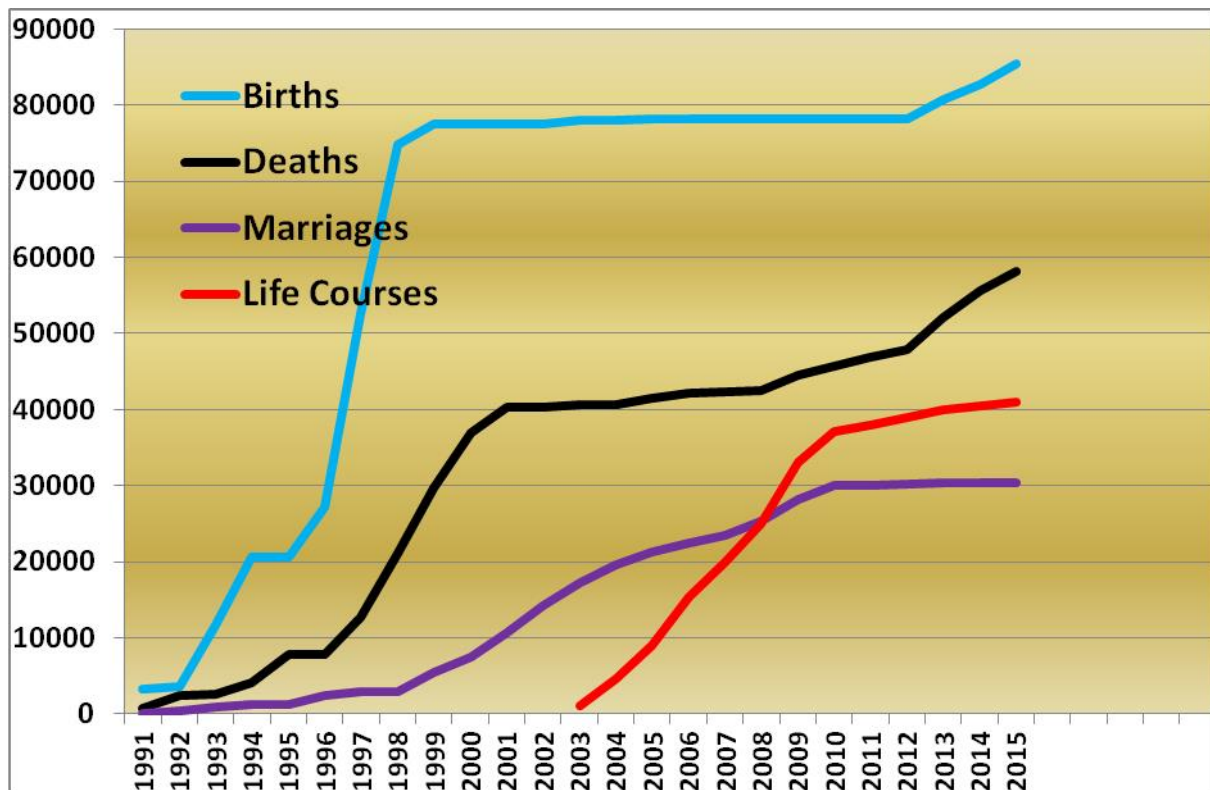
In 2012 the HSN began to complete the sample of the birth certificates of the period 1903-1922. A main part of this sample was entered on the basis of a sample frequency of 0.25% instead of the aimed 0.5%. During 2015 another 2,500 birth certificates were adde to the database. This means that the sample is complete except about 200 from the province of South-Holland. After completing the period 1903-1922 the whole sample will contain 85,500 births.

The maximum number of all sources to be entered for the cases is defined by the number of birth certificates. From figure 1 it is clear that for the life courses the HSN is nearly half way and for the combination of death certificates and personal cards at about two thirds of the number of births. The number of marriage certificates is about 60% of the expected number which is
about 50,000 (not all persons marry or reach a potential marital age).

The fact that the HSN is not yet complete poses a selection problem for each researcher. If and how the data are used depends on the research question and the selection the researcher will make from the dataset, see the following tables 1 and 2 for more detailed information.

Table 1   Number of birth and death certificates and personal cards in HSN dataset by period of birth, 31st of December 2015.

Period	HSN Basic Sample (Number Birth Certif-icates)	Death Certificates and Personal Cards (PK) and Personal Lists (PL)				
		Death Certificates		PK's & PL 's	Death Certificates and PK's & PL's	
		N	% < 16 year	N	N	% Basic Sample**
1812-1862	36,280	17,542	57.7	750	18,292	50.3
1863-1882	16,502	8,732	61.7	5,351	14,083	83.3
1883-1922*	32,572	6,569	79.9	19,228	25,797	79.2
Total 31-12- 2015	85,354	32,843	63.2	25,329	58,172	68.2
Total 31-12- 2014	82,704	32,843	63.2	22,856	55,699	67.3

*   Sampling frequency 0.5% (except small part of South-Holland with 0.25% for the period 1913-1922).
*   The percentages of deaths exclude double counting (of certificates and personal cards

Table 1 presents the database for three periods. We see that for the period 1863-1882 the percentage of found death records is about 83%, For the period 1883-1922 the percentage is about 4% lower, due to the ongoing data entry of birth certificates for the period 1903-1922 and because some persons are still alive. In the early years of the HSN the focus was on the data entry of death certificates of infants and children. This means that these deaths are still overrepresented in the HSN database, although for the two last periods the percentages in the table are exaggerating the situation since the personal cards are not included in this percentage (containing only adult HSN research persons).

Table 2 presents the number of life courses that have been taken in production during the period 2000-2010, all in all 44,252 cases. The HSN used data entry schemes based on a) a distinction in birth period: 1863-1882 and 1883-1922 in which a large part of the sample was prioritized and b) region: the provinces of Utrecht, Zeeland, Friesland and the city of Rotterdam which acted as spearheads. For these areas there was no prioritization in the sampled persons but all of them were followed and completed, and the life courses for the period 1850-1862 were inclied and also the sample size was already set at the necessary 0.5% for the period 1903-1922.

The actual data release comprises a number of 37,173 life courses. Table 2 also presents a bifurcation of the life courses by region and period. Almost 2/3 of the included cases have a complete life course which means that they could be followed from the cradle to the grave or till the year 1940 when the personal card became the only form of population registration. The HSN is still working on the incomplete cases. However, due to emigration, loss of registers

(damage by water or fire), loosing track of persons, quite a lot of these cases will never have a complete recording of their life course.

Table 2   Number of Life Courses by region, date of birth and priority of data entry, HSN Release 2010.01.

Region	Priority	Period of Birth	Total	In release		Complete Life Course	
				N	%	N	%
Spearhead Regions	X	1850-1882	6,208	5,827	93.9	4,179	67.3
Rest of Netherlands	X	1863-1882	6,795	5,608	82.5	4,009	59.0
Rest of Netherlands		1863-1882	5,931	2,159	36.4	1,785	30.1
Spearhead Regions	X	1883-1922	6,528	6,309	96.6	4,805	73.6
Rest of Netherlands	X	1883-1922	14,150	13,185	93.2	10,113	71.5
Rest of Netherlands		1883-1922	4,640	4,085	88.0	3,081	66.4
Total			44,252	37,173	84.0	27,972	63.2

*   Spearhead regions are the provinces of Friesland, Utrecht and Zeeland and the city of Rotterdam. For these regions the life courses from the period of birth 1850-1862 are also included. And for the three provinces the life courses from the oversampling 1903-1922 (from 0.25 to 0.5%) are included as well.

Most of the 7,000 persons who are not included in the release, originate from the birth period 1863-1882. From these 7000 persons 2,500 persons have been collected and entered into the database of which 500 in the report year. So far, a number of 1,000 persons could not be tracked in the registers (mostly because of incomplete registers). The resulting number of 3,000 is in different stages in the process of data collection and data entry. Besides work on the extension of the life course dataset has already started, especially for Amsterdam, the Hague and the provinces of South-Holland and Noord-Brabant for persons from the birth period 1850-1862 (not included in table 2). So, for the life courses the HSN focussed on collecting and data entry. Figure shows the different stages in the production process for all cases from the birth period 1843-1922 which are not included in the release 2010.01. After the collection of data in the region of birth the files are checked, in case the persons had moved to another area, the file moves to the 'mail-system' in which the HSN produce requests for copies of registers filling the gaps that were found in the life courses.

Figure 2   Number HSN Research Persons with no finished life course, per production stage, 1843-1922 (n=27,599).
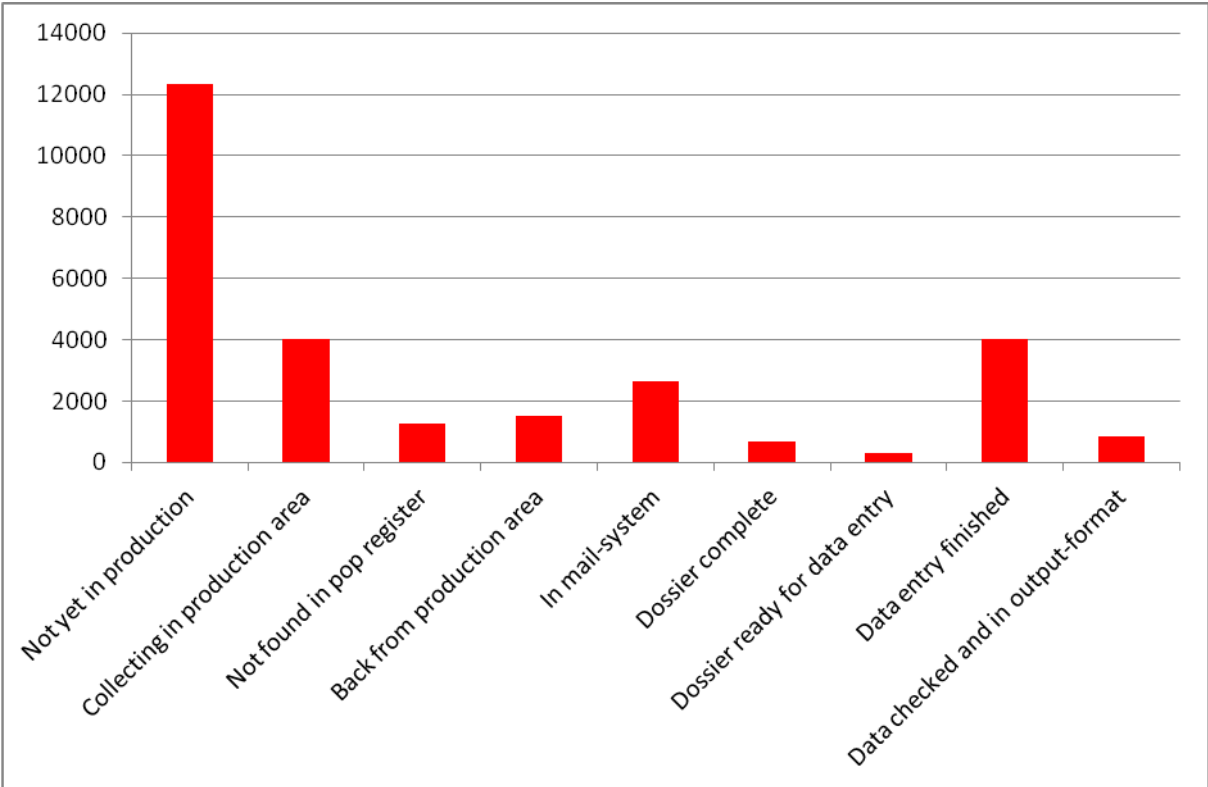
4   Related Datasets.

The HSN functions as an important source for research and as a source for control groups for interpretation of research into specific groups, and besides that the HSN database serves as the basis for collecting new data.

In practice this is realized by:

   a) Designing and maintaining a data structure for use by individual researchers;
   b) Taking the database as a starting point for further research, both by increasing the number of individuals included (oversampling) and by adding supplementary variables for a specific group of research subjects.

Researchers may thus kill two birds with one stone. They can use the data already recorded, as well as the software and expertise developed by the HSN. This expertise is an important byproduct of the data entering of the past ten years. For researchers wishing to use its software and already recorded data, the HSN sets the precondition that new data must be added to the data set, so that these data will eventually become available to other researchers too.

Genes, Germs and Resources
An example of the creation of a related dataset is the project *Genes, Germs and Resource.* The project researches the role of the family and the disease environment in mortality and longevity in the Netherlands, 1812-2015. The project is granted by the Netherlands organization for Scientific Research (Free Competition Humanities and is supervised by Angélique Janssens (Radboud University Nijmegen) and Eline Slagboom (Leiden University Medical Center).

This research project proposes to study the phenomenon of familial influences on early death and exceptional survival in the Netherlands between 1812 and 2015, taking into account the simultaneous effects of *resources, germs and genetic influences*. 'Resources' are defined in socio-economic, and cultural terms; 'germs' refers to the disease environment, and 'genetic influences' refers to an individual's genetically determined predisposition for exceptional survival or the lack thereof. The influence of these factors will be studied through a multi-generational approach in which families are followed over a time span of 200 years.

The goal is to uncover the role of familial influences on survival and the changing interactions between social-structural and biological-genetic factors in mortality and longevity within changing disease environments from the nineteenth and twentieth centuries until today. The project has several innovative aspects, among which the introduction of genetics into the study of historical mortality, as recent advances in human genetics have shown the relevance of the genetic component for longevity and mortality.

The HSN created a dataset named 'Long Lives' which started with the HSN basic set from the birth period 1860-1875. Two groups were distinguished: a) Persons who reached the age of 80 years or older and b) a control group with persons who died before the age of 65. From these persons data about their off-spring were gathered from the population register (second, third and fourth generation).

Geographic and Social Mobility of Female Domestic Servants in Zeeland, 1850-1950
The project on the domestic servants of Zeeland was a joint project with the University of Utrecht and the Inter-Universitary Center for Social Science Theory and Methodology (ICS). The project aimed to reconstruct the life careers of women born between 1835 and 1927 on the islands of Noord- en Zuid-Beveland, Schouwen-Duiveland, Tholen and St. Philipsland. The HSN sample contained 600 of them, 450 of whom had passed the age of fifteen. Sisters were added to the sample, with a maximum of one in each family. This lead to a total sample of approximately 730 women. The researcher, Hilde Bras, focused on the life opportunities of domestic servants compared to other female and male occupations. The research tried to answer the question of how these life opportunities depended on parental environment, migration and the employer's status.

Data were gathered concerning migration, marriage, family situation at the time of birth and at entry into the labour force, income and real estate of the fathers, employers and spouses. The results of this project were presented by Hilde Bras at several conferences, and in a publication called 'Domestic service, migration and the social status of women at marriage. The case of a Dutch sea province, Zeeland, 1820-1935', in *Historical Social Research* 23 (1998).

June 10th, 2002, Hilde Bras presented her thesis *Zeeuwse meiden. Dienen in de levensloop van vrouwen, ca. 1850-1950* at the University of Utrecht.

5   Summary

The Historical Sample of the Netherlands (HSN) aims at constructing life histories as completely as possible for a representative portion of the nineteenth and twentieth century population in The Netherlands. The sample has been drawn from all persons born in The Netherlands between 1812 and 1922. Ultimately, the HSN database will include information at an individual level from about 85,000 persons on subjects like family structure, occupation, birth place, literacy, social network and migration history.

These characteristics make the data set a basic resource for historical research into the areas of demography, sociology, epidemiology, genetics, economy and social geography. The importance of the HSN for the researcher is fourfold:

- The HSN provides a representative dataset which enables research into social developments in the 19th and 20th centuries.
- The HSN provides a control group or groups for researchers to compare with their own research population.
- The HSN is developing the expertise which individual researchers usually cannot acquire in the limited time at their disposal.
- The HSN offers the possibility for researchers to use the existing HSN dataset as a base for their own research projects.

Of course, this cuts both ways. Every researcher who wants to use the infrastructure and data of the HSN must agree that in return he or she will deliver his or her data to the central database, in accordance with the formal structure of this database. In this way the HSN has developed into a data centre that functions as a centre for quantitative research on life courses