Reconstructing the Extended Kin-Network in the Netherlands with Genealogical Data: Methods, Problems, and Results

Wendy Post; Frans Van Poppel; Evert Van Imhoff; Ellen Kruse


Stable URL:
http://links.jstor.org/sici?sici=0032-4728%28199711%2951%3A3%3C263%3ARTEKJIT%3E2.0.CO%3B2-Y

Population Studies is currently published by Population Investigation Committee.

Your use of the JSTOR archive indicates your acceptance of JSTOR’s Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR’s Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/pic.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org/
Fri Jun 2 10:54:44 2006
Reconstructing the extended kin-network in the Netherlands with genealogical data: Methods, problems, and results

WENDY POST, FRANS VAN POPPEL, EVERT VAN IMHOFF AND ELLEN KRUSE

I. INTRODUCTION

The demographic history of kinship is a subject that, despite all progress in the empirical methods of historical demography, is still in an underdeveloped stage (Smith and Oeppen 1993, p. 280). This is largely owing to the lack of sources of evidence on the kinship networks of individuals over the entirety of their lives (Plakans 1984). This is not only a problem for historians, but also for contemporary and prognostic kinship research. The recent interest in kinship relates to the fact that demographic changes will reduce the capacity of the kin network to look after their elderly relatives (Dykstra and Knipscheer 1995). Insight into the availability of support for the elderly requires precise information on the actual size and structure of the kin network. Heran (1982) distinguishes two methods of collecting such information: surveys and the construction of theoretical models.

Until now, surveys have been few in number and usually limited to a very restricted group of kin. They often relate to a few hundred respondents only and thus might not be considered representative of the population as a whole. In France, some more extended surveys were held in the 1970s and 1990s (Gokulp 1978; Bonvalet et al. 1993). An important advantage of surveys is that they deal with a real population. However, they give only cross-sectional information on kin networks, thus precluding longitudinal analyses. Adding retrospective questions does not really solve this problem: the resulting data refer to a selected group of individuals only (survivors), and memory problems may seriously bias the results.

To solve the problems posed by the use of survey data on contemporary kin networks, demographers have increasingly turned to theoretical models. For obvious reasons this is also the preferred solution when the focus is on the future size and structure of the kin network. Historians have recently adopted this approach to overcome the lack of historical data.

Existing theoretical models vary strongly in their complexity and mathematical methods (for an overview see Smith and Oeppen 1993, pp. 281–2).

Basically, what a theoretical model does is to describe kin relations as the outcome of demographic events. Once demographic behaviour is specified, the resulting kin relations can be studied, either analytically (e.g. Goodman et al. 1974; Pullum 1982) or through simulation. Recently, microsimulation models in particular have become popular (e.g. Galler 1990; Nelissen 1994). The best-known model among historians is the CAMSIM model developed by Smith and Oeppen (1993).

Model kin counts are not usually legitimate substitutes for empirical kin counts (Smith 1987, pp. 249–250). Models are abstractions from reality, and the degree of realism in the model assumptions can always be debated. In particular, because of scanty data, supplying complete time series of input parameters raises huge and often insurmountable problems, particularly for the pre-1900 period (Watkins et al. 1987). As a consequence, model builders are forced to introduce simplifying assumptions about the underlying demographic processes.

Several authors have suggested that a solution to the problem of lack of historical data on the demographic experience of kin groups as a whole might be found in genealogical data, deduced from the entries in a parish or civil register (Plakans 1984; Ruggles 1993; Smith and Oeppen 1993; for an application, see Zhao Zhongwei 1994). A record of a genealogical nature includes precisely the information at the individual level that makes it possible to reconstruct the individual’s kinship network: the entry of an individual into a genealogical configuration immediately connects him or her with an entire network of relatives of various kinds. Because such a reconstruction is based on thousands of records of vital events, it has a potential for a very elaborate genealogical configuration that incorporates large stretches of historical time. Both the continuities and discontinuities in the size and composition of that network for a particular group or region might thus be established.

To test this supposition, it was decided to use genealogical data for the study of the historical
development of kinship networks in the Netherlands during the period 1830–1990. This offered the opportunity to assess the effect of the first and the second demographic transition on the numbers of kin of various types at various points in an individual’s life. By including also the most recent past, it also became possible to study the usefulness of genealogical sources as a substitute for survey information on contemporary kinship.

The organization of this paper is as follows. Section 2 discusses some general problems in using genealogical data, particularly the issues of representativeness, accuracy, and completeness. The genealogical sources used for the present study are described in section 3, and the quality of the data is investigated in section 4. As will become apparent, one particular problem with the genealogical data is the fact that in many cases the year of death is missing. To remedy this deficiency, the missing death years have been estimated; a description of the estimation procedure used is given in section 5. The imputed genealogical database has been used to analyse the development in the structure of Dutch kinship patterns over the period 1830–1990. The results are presented in section 6. A summary of the main findings is given in section 7.

II. PROBLEMS IN USING GENEALOGICAL DATA: GENERALITIES

The kind of genealogical information needed for the reconstruction of the demographic regime and the kin network are patrilineage tables, also called descendant genealogies, or family histories. Until recently, most investigators were doubtful about the adequacy of genealogies as a source for demographic research. The major difficulties arising from the use of genealogies are the various kinds of under-registration of individuals and events, and the lack of representatives of the study population (Hollingsworth 1969; Henry 1956; Dupâquier 1993; Smith and Oeppen 1993, pp. 280–281).

The representativeness problem has several aspects. The first relates to time and space. In demographic research, a population is usually defined as a group of people living in a given place at a given time. In research using genealogies, reference to a population is much less unambiguous. Most genealogical reconstructions have no locational specificity. Each individual is entered regardless of where he or she was born, married or died. Even if genealogies had a specific regional origin, the characteristics that originally defined the population under observation might be diluted over time (Norton 1980; Hollingsworth 1976).

The representativeness problem is also related to the small number of families normally included (Jetté and Charbonneau 1984). In most studies, genealogies include at a maximum only some hundreds of marriages. Biological chance alone and the specific historical context then play a large role and comparisons with or generalizations to a larger population are rather dangerous.

Written genealogies have historically tended to be limited to the privileged or powerful members of society. This might lead to selection: the probability that a family history will be constructed increases with the fertility of the family and decreases with its mortality (Henry 1956; Fogel 1993). However, during recent decades this problem has become less serious, as large collections of genealogies of ordinary people have become available (Fogel 1993). What is more, external sources like census records can be used to evaluate the representativeness of the genealogical database (see Skolnick et al. 1984; Rallu 1992).

The extent to which one is able to reconstruct in an accurate way the complete kinship network of a population depends on the degree to which all relevant branches of the descendant genealogy have been traced, on the completeness of the demographic information about the members of the descendant genealogy, and on the extent to which this information is accurate.

The completeness of the genealogies and their accuracy are strongly related to each other. They depend both on the sources available for the compilation of the genealogies and the quality, zeal, and available time of the genealogist. However, the most important factors are the degree to which all relevant branches of the descendant genealogy have been traced and the completeness of the demographic information about the members of the genealogy. As genealogists usually are less interested in events which had no effect on the development of the family history and have more difficulty in tracing events which took place outside the family circle (Jetté and Charbonneau 1984), several very specific kinds of under-registration of individuals and of events take place (Zhao Zhongwei 1994).

As there is a close attachment to the notion of the ‘family line’ which is continued from generation to generation and such a succession is widely supposed to be carried out only through male descendants, female members of a family are frequently excluded from the genealogy. This also applies to people who die young. Most genealogies start with a person who lived long enough to see his or her sons or grandsons and therefore had a better
chance of being recorded; those who belonged to early generations but died young were most likely to be excluded. Thus, the mortality which was actually experienced by these early generations tends to be under-represented in the genealogies. The last few generations recorded in a genealogy often exclude information on date of death of persons who were still alive at the time the genealogy was compiled. Persons who married into the family are frequently excluded from the genealogy (see Rollet 1973); as a consequence, the kin ‘in law’ is typically missing. Divorces and remarriages also fall into this category of event-related omissions. Finally, almost no genealogy, especially those which cover long periods, includes everyone who descended from a common ancestor. Family division, migration, or similar events might mean that the family lost track of those who had left and could not obtain any information on their descendants. These branches are therefore not included in the genealogy (branch-related omissions).

For the reconstruction of the size and composition of the kin network, event-related and branch-related under-registration are particularly important. In numerous descendant genealogies, the descent of no more than a limited number of children (branches) of each parent has been traced. Each time that a genealogist did not wish or could not follow the individuals or couples who migrated outside the region, the genealogy is deprived of a part of its ramifications. Information about the next of kin (parents, grandparents, brothers and sisters) of spouses is even less common. Omissions might also relate to events or dates of events. This category of omissions is sometimes directly visible, e.g. when the date of birth or death is not mentioned but sometimes it is not, e.g. when no marriage is mentioned.

At first sight, the best solution seems to be to take only those genealogies that are based on impeccable records. However, if all genealogies with partial or only approximate data are eliminated from consideration, the number of cases available for study may be sharply reduced. In addition, such a decision would seriously limit the range of variation, and might bias the results of the analysis if there is a systematic relationship between the demographic experience of various subgroups and the relative completeness of their records (Knodel and Shorter 1976).

III. DATA COLLECTION AND DATA PROBLEMS

The source material available is an important determinant of the quality of genealogies. Dutch genealogies are mainly based on four kinds of sources: the vital registration system; the baptism, burial, and marriage registers; the population registers; and the personal cards.

For the period before 1811, data are available from parish registers only. The extensiveness and accuracy of information available in these baptism, marriage, and burial registers is rather deficient because there was no generally prescribed system for registering these events; all religious groups kept separate registers, and there was no powerful central authority.

Civil registration of births, deaths, and marriages in the Netherlands as a whole only started in 1811. In 1849, the scope for identifying people in the past was greatly increased by the introduction of the continuous municipal population register. Its main advantage is that it contains information on the time and destination of migration of all inhabitants. In the 1930s, the personal card was introduced, which made the individual person the registration unit in all municipalities.

In general, the quality of information available to the genealogist might be considered good enough to allow the reconstruction of the whole family history. However, the odds of finding complete information on kin members greatly increase the more one approaches the twentieth century. On the other hand, for the most recent period, restrictions on access to the vital registration system might result in under-representation of events (marriages, deaths, births etc.) which took place only recently.

The limitations described might have had consequences for the representativeness of our genealogical database. Persons belonging to the first few generations who died at lower ages are more likely to have been excluded from the genealogies. Since in the period 1730–1810 the amount of information available increased considerably, the probability of having information on the kin network of persons alive in 1830 or later years is higher among younger age groups. The resulting population in the first decades of the nineteenth century is thus younger than it actually was.

Given restricted amounts of time and money, it was decided to include in this study only those genealogies which were already available in machine-readable form. Various procedures were used to contact genealogists who had entered a descendant genealogy into a computer program. As a result, genealogies were received from a great
diversity of regions. The degree to which the population derived from these genealogies can be regarded as typical of the Dutch population will be dealt with in a following section. For the moment, it is assumed that the genealogies refer to the population living in the Netherlands during the period 1830–1990. For the final analysis, 124 descendent genealogies were used, drawn up by 54 different genealogists. They contained information on a total of 162,354 persons.

The internal consistency of the descendant genealogies was checked. If inconsistencies were found, values were either set to missing or corrected manually. In making such corrections, use was made of ‘best guesses’, and sometimes of additional ‘comments’ that were included in the computer files. Our controls could not detect all mistakes in the database, but the end result seemed fairly reliable. All computer files were finally combined into one big database. Possible kinship ties between different descendant genealogies were not taken into consideration; each descendant genealogy was taken as a separate family.

In order to be able to reconstruct kinship networks, information is needed about three kinship ties for each individual person: about parents, children, and partners. If these three types of relations are known for all individuals, all other kinship ties can, in principle, be deduced. Since we want to know the size and structure of the kinship network at any point in time, we must also know when each person in the kinship network was born, when he or she died, married, and when the marriage broke up. These data will be designated the demographic characteristics of the members of a descendant genealogy.

For each individual, the database contains a record with the following variables: an identification number, sex, year of birth, year of death, most recent marital status, and the number of marriages contracted. For a maximum of three marriages, the record also includes the identification number of the partner, the date of marriage, and the date and cause of marriage dissolution. Finally, the record includes the identification numbers of the parents and of all children (up to a maximum of ten) of the individual concerned. If no information was available regarding a certain variable, the value ‘zero’ was entered.

<table>
<thead>
<tr>
<th>Period of birth</th>
<th>Males</th>
<th>Females</th>
<th>Unknown death-year (per 100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1730–39</td>
<td>409</td>
<td>388</td>
<td>53.5</td>
</tr>
<tr>
<td>1740–49</td>
<td>459</td>
<td>401</td>
<td>50.3</td>
</tr>
<tr>
<td>1750–59</td>
<td>556</td>
<td>531</td>
<td>48.6</td>
</tr>
<tr>
<td>1760–69</td>
<td>586</td>
<td>555</td>
<td>47.3</td>
</tr>
<tr>
<td>1770–79</td>
<td>658</td>
<td>672</td>
<td>44.7</td>
</tr>
<tr>
<td>1780–89</td>
<td>825</td>
<td>833</td>
<td>52.7</td>
</tr>
<tr>
<td>1790–99</td>
<td>1174</td>
<td>1097</td>
<td>51.4</td>
</tr>
<tr>
<td>1800–09</td>
<td>1272</td>
<td>1280</td>
<td>45.6</td>
</tr>
<tr>
<td>1810–19</td>
<td>1353</td>
<td>1324</td>
<td>43.9</td>
</tr>
<tr>
<td>1820–29</td>
<td>1750</td>
<td>1759</td>
<td>47.9</td>
</tr>
<tr>
<td>1830–39</td>
<td>1674</td>
<td>1612</td>
<td>45.6</td>
</tr>
<tr>
<td>1840–49</td>
<td>1621</td>
<td>1544</td>
<td>43.4</td>
</tr>
<tr>
<td>1850–59</td>
<td>1689</td>
<td>1728</td>
<td>40.7</td>
</tr>
<tr>
<td>1860–69</td>
<td>1977</td>
<td>1940</td>
<td>40.6</td>
</tr>
<tr>
<td>1870–79</td>
<td>2190</td>
<td>2149</td>
<td>38.4</td>
</tr>
<tr>
<td>1880–89</td>
<td>2508</td>
<td>2347</td>
<td>36.2</td>
</tr>
<tr>
<td>1890–99</td>
<td>2639</td>
<td>2549</td>
<td>35.7</td>
</tr>
<tr>
<td>1900–09</td>
<td>2475</td>
<td>2623</td>
<td>43.2</td>
</tr>
<tr>
<td>1910–19</td>
<td>2575</td>
<td>2598</td>
<td>58.5</td>
</tr>
<tr>
<td>1920–29</td>
<td>2754</td>
<td>2712</td>
<td>78.1</td>
</tr>
<tr>
<td>1930–39</td>
<td>2301</td>
<td>2245</td>
<td>90.1</td>
</tr>
<tr>
<td>1940–49</td>
<td>2550</td>
<td>2456</td>
<td>92.8</td>
</tr>
<tr>
<td>1950–59</td>
<td>2332</td>
<td>2315</td>
<td>97.0</td>
</tr>
<tr>
<td>1960–69</td>
<td>2026</td>
<td>2015</td>
<td>96.6</td>
</tr>
<tr>
<td>1970–79</td>
<td>1313</td>
<td>1247</td>
<td>98.4</td>
</tr>
<tr>
<td>1980–89</td>
<td>1003</td>
<td>899</td>
<td>98.3</td>
</tr>
</tbody>
</table>

IV. THE COMPLETENESS AND REPRESENTATIVENESS OF THE DATABASE

The practical possibilities of the database are mainly determined by the extent to which all branches of the descendant genealogies have been traced and by the completeness of the demographic data of the members of the descendant genealogies. In this section, the impact of both these phenomena is determined and solutions for the problems which arise are suggested.

4.1. Demographic information

Obviously, year of birth, year of death, and sex are essential pieces of information in the reconstruction of the kin network. Table 1 gives an overview of the numbers of people in the database – decomposed by period of birth – and the percentages in each cohort for which the year of death is not known. It is disappointing that the year of birth of 71,353 people (43.9 per cent of the overall total) is not known. As a result, data could be used for 84,488 people only. The year of death is unknown for about 44.4 per cent of that part of the group for whom this should, in the nature of things, have been known, namely those born prior to 1890. The number of people for whom the year of death is

266
unknown varies by between 40 and 60 per cent per cohort for females and by between 36 and 53 per cent per cohort for males. It slightly increases from the oldest to the more recent cohorts. Surprisingly enough, percentages unknown for females (45.3) are hardly higher than those for males (43.7).

To determine whether the database was representative of the total population of the Netherlands, its composition by age, sex, and marital status was compared with that of the total population. In addition, some vital statistics were calculated for the descendant genealogies and compared with annual vital statistics. A $\chi^2$-test was carried out for each calendar year in order to determine whether the age distribution of the database population differed from that of the total population. The highest age category (85+) and the youngest age category (0–4 years) made the largest contribution to the $\chi^2$ statistic. Since the year of death is missing in a large number of cases, one would indeed expect an over-representation of the older age categories. It was also found that the age distribution of men in the database deviated more strongly than the age distribution of women. Deviations increased between 1850 and 1920, and became smaller after 1920. Thus, the influence of missing years of death was smaller for recent periods, and this is reflected in a declining contribution of the highest age categories to the $\chi^2$ statistic.

Calculations showed that the distribution of the research population by marital status also deviated from that of the total Dutch population. Divorced people and widowed people were particularly under-represented. Apparently, the data on changes in marital status, especially the available data on divorce and widowhood, are too incomplete to be included in the analysis.

On the basis of the above findings one may conclude that the lack of data regarding the year of death constitutes a problem which seriously restricts the representativeness of the database for the entire period 1850–1990 (and, one must assume, also for the period 1830–50). It is possible, however, to find a satisfactory solution to this problem by estimating, from life tables, the year of death for those people for whom the date of birth is given but the year of death is not known. Section 5 deals with the procedure in more detail.

4.2. Missing branches

Another source of incompleteness of the database lies in the fact that the genealogists merely traced the descendants of some of the children in a particularly descendant genealogy and did not trace all branches. This type of incompleteness would be reflected in the facts that a large number of descendents who actually married appear to have remained unmarried, and that a large number of marriages appear to have remained childless whereas children were, in fact, born. In short, the database would contain too many zeros for the number of children. In the following, we will use the terms missing zeros and structural zeros for the number of children. A missing zero is a zero in the database that results from the fact that the branch was not traced (i.e. a person did in fact have children but they were not included by the genealogist). A structural zero is a zero that corresponds to the actual situation (i.e. the person really remained childless).

Comparing the expected percentage of structural zeros with the total percentage childless people might give an idea of the number of descendant genealogies that had not been fully traced – the missing zeros. Although the trends in the percentage childless and in the distribution by parity in the research population coincide more or less with trends in the Dutch population, the actual levels of these fertility indicators deviate considerably. The percentage childless in the descendant genealogies appears to be fairly stable at between 53 and 60 per cent. Compared with the estimated incidence of childlessness in the total population, this percentage is much too high. The percentage of missing zeros in the period 1830–1950 is estimated to be about 30 per cent, and about 40 per cent in the years 1960–90. Thus, around one out of four persons who were classified as childless in the period before 1900 were incorrectly so classified. After 1970 this situation became even worse, owing to the fact that in the more recent period it became more difficult to update marriage and childbirth data for family members.

Another consequence of the failure to trace the descent of all branches in a descendant genealogy is that one-child families are over-represented, especially at the top of a family tree. This is linked to the fact that compilers of genealogies are frequently interested in direct lines of descent only. In the rest of the family tree, all the children are usually mentioned, but their descent is not always traced. This bias affects the percentage of people who have had only one child. Indeed this percentage decreased after 1830.

A final aspect of missing branches relates to family size. The larger the number of children in a given family, the less likely it seems that the descent
of all the children will be traced. Since research has shown that there is a relationship between the size of ego’s parents’ family and the size of ego’s own family (e.g. Ruggles 1993), the number of people with large families may well be under-represented in the descendant genealogies.3

4.3. Surmounting the problems of under-registration

It is clear from the preceding sections that calculations of kinship ties based on the fertility pattern found in this research population will not give an accurate picture of the size and composition of the kin network. In principle, this problem cannot be solved with the available database. The necessary data are simply lacking. However, by applying a more restricted definition of the composition of the kinship network, and using only those pieces of information that are most reliable, a practical solution can be found.

With respect to the missing branches, our practical solution is to exclude from the analysis individuals without children (in the downward direction) and individuals with unknown parents (in the upward direction). In effect, all zeros are discarded. Since every person has two parents, all missing branches in the upward direction are missing zeros by definition. Since some people remain childless, all missing branches in the downward direction contain a mixture of missing and structural zeros on the number of children. Thus, the analysis is limited to individuals with at least one child. Although the number of children will still be under-estimated by the genealogical data, the most important source of under-estimation has been neutralized. For the various types of kin, discarding zero parents and zero children implies the following restrictions.

- children: only egos with at least one child;
- grandchildren: only egos with at least one grandchild;
- siblings: only egos with at least one known parent;
- nephews/nieces: only egos with at least one nephew/niece (at least one non-childless sibling);
- aunts/uncles: only egos with at least one known grandparent;
- cousins: only egos with at least one cousin (i.e. at least one non-childless aunt/uncle).

With respect to event-related under-registration, clearly the most unreliable data are those on marital status. This has two main consequences for the analysis of kinship patterns. First, for many marriages, information on the spouse or the spouse’s kin is missing. For this reason, we will consider consanguineous kin only. Kin relations that run via the spouse, i.e. the kin ‘in law’, will be ignored. Second, information on remarriage is of a dubious nature. As a result, the allocation of children to the marriage out of which they were born is unreliable too. For instance, in many cases of remarriage, the children are registered under the ‘wrong’ parent. Thus, since no clear distinction can be made between blood ties, on the one hand, and step- and half-ties, on the other hand, the analyses include both step-ties and half-ties.

When interpreting the results of the analysis, to be presented in section 6, these restrictions should be kept in mind. Despite these restrictions, the numbers of kin continue to be under-estimated, owing to the gaps inherent in the genealogical database.

V. IMPUTATION OF THE YEAR OF DEATH

In section 4.1 it was concluded that the lack of data on the year of death constituted a big problem. Therefore, it was decided to impute (i.e. to estimate) the year of death for those individuals in the database for whom the year of birth was given but not the year of death. The mortality data used for this imputation procedure have been taken from historical cohort life tables for the Netherlands, covering the period 1850–1990 (Tabeau et al. 1994).

Suppose we have an individual of sex z born in year x with unknown year of death. If we knew the appropriate probability distribution of years of death for sex z and birth cohort x, we could draw a random number and substitute the corresponding death year into the database. On average, such a procedure would replicate the actual mortality experience.

There are several problems in using this imputation scheme. Most important is the fact that whether or not genealogical information on the year of death is available is dependent on the age at which people die. Persons who die before marriageable (child-bearing) age are of only limited interest to the genealogist, since they do not produce offspring. Also, one may assume that it is more difficult to trace the descent of people who die young. The selective nature of year-of-death information is properly taken into account if the population life table is interpreted as a mixture of the known mortality pattern of those with death-
year information and the unknown mortality pattern of those without it. Since the weights of the populations with and without death-year information are known from the genealogical database, the probability distribution of death-years for the selective group of individuals without death-year information can be reconstructed.

Secondly, to reduce the impact of random sample fluctuations, the number of observations for each cell (combination of sex and birth cohort) must be made sufficiently large. All cohorts born prior to 1850 have been taken together; the remaining cohorts were grouped into 10-year intervals (1850–59, 1860–69, etc.) and the population life tables were averaged accordingly. Finally, when applying the probability distribution to impute the year of death, the result could lead to inconsistencies in cases where other dates are known about a given person (birth of children, marriage). To surmount this problem, first a minimum calendar year which a person is known to have reached was determined; then, given this minimum year of death, the subsequent random year of death was drawn from the appropriate tail of the probability distribution.

The representativeness of the resulting imputed database was assessed using criteria similar to those used in section 4 for the original database. From this comparison, it can be concluded that the data in the imputed database are more realistic than those in the original one. The age structure of the imputed database coincides more strongly with the age structure of the total population, and demographic indicators like the crude birth rate are more realistic. For this reason, the analysis in section 6 will be restricted to data from the imputed database.

VI. THE KINSHIP NETWORK: RESULTS

In this section, results will be presented for the imputed genealogical database, covering the period 1830–1990, for six types of kin: children; grandchildren; siblings; nephews and nieces; aunts and uncles; and cousins. We concentrate on the average number of living kin. In addition, we should stress once more the specific definitions that we used for each of these types of kin, which were discussed in section 4.3 above.

Figure 1 illustrates the average number of living kin over the period 1830–1990 for six types of kin, averaged over the whole population. Figures 2–7 give the same averages for four age categories: persons under 20, persons between 20 and 40 years old, between 40 and 60 years, and aged 60 and over. These age groups will be referred to as the youngest generation (or young people), the young adults, the intermediate generation, and the elderly, respectively.

6.1. Children

The development of the average number of living children (irrespective of their age) born to people who had at least one child (Figure 1) is characterized by an increase until about 1920 (from about 2.2 in 1830 to 2.7 in 1920) and later by a decrease to 2.4 in 1990. Thus the strong decline in the number of children, which one might have expected on the basis of the fertility decline which set in during the last quarter of the nineteenth century, did not have major consequences for the average number of living children for individuals with at least one child. There are several reasons for this. First, mortality trends played a neutralising role: extremely high (child) mortality kept the averages low at the beginning of the period studied. The decline in child mortality after about 1880 may have triggered the slight increase until 1920. Second, large families are under-represented in the genealogies; this downward bias was more important during the period in which ‘large’ families were common than in recent years. Third, to the extent that the fertility decline was caused by an increase in childlessness (structural zeros), there was no impact on the average number of children for individuals with at least one child (recall that our definition of kin excludes all zeros). However, for the same reason, one may not attribute the limited variation over time to trends in the incidence of missing zeros.

Of course, the data in Figure 1 are influenced by the age structure of the population for which the family size has been calculated. With time, the average age of this population rises, which increases the probability that their children are no longer alive. This problem is solved in Figure 2, which gives the development by age group. Until about 1920 the intermediate generation has the largest number of living children; the elderly on average have around 0.25 children fewer. In both groups, the number of children alive increases over time to values of 3.0 and higher in the first decades of the twentieth century. Only after 1950 does the average number of children among the elderly and the intermediate generation start to decrease. The number of children among persons aged 20–40 is much more stable over time and much lower than among older persons. Of course, this group is still of child bearing age. The number of children among the 20–40 year olds also shows an initial
increase, until 1890; after this date the average gradually declines to 2.1 in 1990. It is indeed remarkable that in 1990, for all age groups, the average numbers of children are as high as, or even higher than, average numbers in the early nineteenth century.
Figure 3. The average number of living grandchildren for persons with at least one grandchild, by age group: Netherlands, 1830–1990.

Figure 4. The average number of living siblings, by age group: Netherlands, 1830–1990.
6.2. Grandchildren

As mentioned earlier, the offspring of about 30–40 per cent of the people in the descendant genealogies has not been traced. Because of these missing zeros, the average number of living grandchildren (Figure 1) is under-estimated in the genealogical database. The average number of grandchildren among people with at least one grandchild clearly increases until about 1930, and slightly declines after 1950. Between 1920 and 1950, the average number of grandchildren is between 4.5 and 4.7.

Figure 3 gives the averages for two age categories. The average for the 40–60 age group is, of course, substantially lower than the average for the elderly. The number of grandchildren in this intermediate generation is remarkably stable. For the group of elderly, the average number of grandchildren has increased considerably, reaching the highest value in 1930. Subsequently the trend is downwards, until the present day; however, the number of living grandchildren in 1990 is still higher than it was in the nineteenth century.

6.3. Siblings

Figure 1 gives the development of the average number of brothers/sisters. The number of siblings increases steeply from around 2.1 in 1860 to 3.0 in 1940; after 1960 it quickly decreases to 2.5 in 1990, still higher than in the nineteenth century.

The average number of living siblings differs considerably between the four age categories (Figure 4). The trends found among the elderly are particularly interesting, with an almost constant increase since 1830 from about 0.7 to 2.5 living brothers and sisters. In contrast, the average number of siblings among the youngest age groups increases only until 1900 and drops sharply after 1920, from around 3.2 to only 1.5 in 1990. Similarly, the intermediate generation shows first a steep increase until the 1970s, and then a decrease. Persons aged 20–40 also see their number of siblings increase; the decrease sets in around 1950. Until 1990, the largest numbers of siblings are found in the age groups 0–20 and 20–40. From around the 1960s, the intermediate and the elderly generations have larger numbers of siblings.

It is clear from Figure 4 that all four age groups show a comparable development over time, but reach the highest numbers of siblings in different time periods: young people reach their maximum around 1900, young adults in 1920, the intermediate generation in 1960, and the elderly in 1990. This sequence of maxima reflects the changing fertility trends over time, which reach each age group at a different moment in their reproductive life.

6.4. Nephews and nieces

The average number of nephews and nieces (the children of ego’s brothers/sisters), for people with at least one, show a continuous increase from 1830 onwards (Figure 1). It reaches a maximum of 7.5 in 1960, compared with no more than 5.0 in 1830. A decline sets in after 1970, lasting until 1990, in which year the average number of nephews/nieces stands at 6.5, still clearly above the values found in the nineteenth and early twentieth century. It should be stressed that the number of nephews and nieces is under-estimated by the genealogies because of missing branches. In addition, this underestimation is probably most serious for the earliest generations.

Decomposition by age group (Figure 5) again shows large differences. The average number of nephews/nieces for people aged less than 20 fluctuates and does not show a clear trend. Persons aged 20–40 have about 2.0 nephews/nieces more than the youngest age group; stability is also characteristic of this age group, until 1970; after this year, a decrease takes place. The intermediate and oldest age groups have the largest number of this type of kin, the differences between the two being only slight. Since the beginning of the nineteenth century, the number of nephews and nieces increases strongly in both age groups, until around 1940 for the age group 40–60 and until the 1970s for the elderly.

6.5. Aunts and uncles

Since people with all grandparents missing were excluded, the data on aunts and uncles relate to only a small part of the research population. Within this small part, information on both lines of grandparents was available for about 30 per cent of the individuals in the nineteenth century, and 15 per cent in the twentieth century. One might have expected this to lead to a decrease in the average number of aunts and uncles over time but in fact it increases continuously (Figure 1), reaching its maximum in the period 1930–1960. Thus the increase in the period until 1930 must have been much stronger than is depicted by the data.

Decomposition of the research population by age group (Figure 6) yields about the same pattern as that for the total population, albeit with
Figure 5. The average number of living nieces and nephews for persons with at least one niece or nephew, by age group: Netherlands, 1830–1990.

Figure 6. The average number of living aunts and uncles, by age group: Netherlands, 1830–1990.
considerable differences between the age groups. The largest number is present among the youngest age group: it increases from about 2.5 in 1830 to around 4.5 in the period 1920–1970. Thereafter, the number of uncles and aunts decreases strongly. The increase is even much stronger among the young adult group: from about 1.3 to 4.0 in 1990. Since the 1940s, the number hardly increases any further. The intermediate generation has a much smaller number of uncles and aunts alive, although in this age group too an increase takes place from around 1890 to 1960. The elderly have hardly any uncles or aunts alive, although in recent decades a slight increase in their number takes place.

6.6. Cousins

The average number of (first) cousins is also heavily influenced by the fact that in about 35 to 50 per cent of cases, at best, only the paternal or the maternal cousins were known. Here too, the average was calculated for those people who had at least one cousin. On the one hand, this procedure slightly over-estimates the average number of cousins, since people with structural zeros are not included; on the other hand, there is under-estimation, since the descent of some aunts/uncles is not included. Thus, the numbers found have an indicative value only; the trends over time are more important than the absolute values. The average number of cousins increases until about 1950 (Figure 1). A slight decline sets in during the last two decades, but even in 1990 the members of the research population have a larger number of living cousins than the nineteenth-century population.

The averages for the youngest, the 20–40 age group, and the intermediate generation do not deviate much from each other until 1920 (Figure 7). After 1920, the trends for these age groups start to drift apart. The number of living cousins for the youngest age group remains rather stable and decreases strongly from 1960 onwards. The average for the age group 20–40 increases until 1930, then remains stable, and finally decreases after 1980. The intermediate generation has an increase lasting until the present day. The oldest generation also experiences a steep increase in the number of living cousins: from about six in the first half of the nineteenth century to nine in 1990. Currently, the elderly have more living cousins than the youngest generation.

6.7. Parents

So far, one type of kin that each person must have had has not been discussed: the parents of ego. In contrast to the foregoing, it is not the average numbers of kin that are presented here but the percentages of people who have a given number of parents alive.
Figure 8. The percentage of people with both parents deceased, by age group: Netherlands, 1830–1990.

Figure 9. The percentage of people with one parent deceased, by age group: Netherlands, 1830–1990.

Figure 8 shows that the proportion of people aged less than 20 who have no parent alive is already extremely low throughout the nineteenth century and decreases even more after 1890. Among people aged 20–40, 20 per cent are orphaned, and the percentage decreases further after 1900 to reach
five per cent in 1990. The majority of the intermediate generation (around 70 per cent) is parentless during the period 1830–1910; from that time on, a decrease takes place, and in 1990 only 40 per cent of people in this age group are full orphans. Among the elderly, orphanhood is the rule; from 1930 onwards, the proportion of full orphans among the elderly decreases to 95 per cent.

Figure 9 shows that the percentage of people who have only one parent alive varies between 0 and 40 per cent according to the age group of ego. Among the young and the young adults, a gradual decrease of this percentage takes place from 1870 onwards; at the same time, the proportion of persons with both parents still alive increases. Among the intermediate and the older generation, the percentage with one parent alive increases. The differences between both groups are nonetheless enormous. The percentage of half-orphans increases among the intermediate generation from 25 per cent to 40 per cent in 1990; among the elderly, the increase starts thirty years later (in 1940) and reaches a value of only five per cent in 1990.

6.8. Grandparents

Information on both pairs of grandparents of ego was available for only 10–15 per cent of the research population. Since the results for this last group are not considered representative, they will not be discussed here. Use will be made of information on only those people for whom data on the grandparents on the father’s side or on the grandparents on the mother’s side were available. For both sides the percentage of people with living grandparents shows only very small changes during the period 1830–1990 for the population as a whole. Again more interesting figures are revealed when the four age groups are distinguished.

People aged 20 or less show a clear increase in the number of living grandparents: from 12 to 42 per cent for those on the father’s side, from 18 to 40 for those on the mother’s side. This increase starts only in the 1920s when, in the age group 20–40, it is very rare for both grandparents to be still alive. This percentage increases during the twentieth century, reaching seven per cent in 1990 for the grandparents on the mother’s side. In older age groups, one or both grandparents are almost never still alive.

If one assumes that the mortality pattern on the paternal side is independent of that on the maternal side, one can estimate what proportion of people had four, three, two, one or not a single grandparent alive. For the population as a whole, the ageing process largely determines the outcomes. As a result of this process, the percentage with only one
grandparent alive decreases, and the percentage with no grandparents alive increases from 54 per cent in 1830 to 59 per cent in 1990. For people aged less than 20, percentages are given in Figure 10. This shows that, whereas in 1830 only two per cent of the younger age group have all four grandparents alive, the figure increases to 20 in 1990. The percentage of people aged less than 20 with three grandparents alive increases at the same time from 9 to 33 per cent. Whereas in 1830 almost 28 per cent of the youngest age group have not a single grandparent alive, this percentage decreases to three in 1990.

VII. CONCLUSIONS

There are two main problems in using existing genealogical data to study the development of the kinship network in the past: the year of death is missing for a sizeable part of the research population, and the information on all relevant branches of the genealogies and the demographic information on the members of the genealogies are far from complete. However, if the effects of the various types of under-registration are carefully assessed, genealogies can still provide valuable information for our understanding of kinship patterns during a long period of our history. For this reason, an attempt was made to find practical solutions for both problems. A mixed estimation procedure was used to impute the missing years of death. The second problem was overcome by making several modifications to the definition of kin, the most important being to exclude individuals without children from the analysis. As a consequence of these practical choices, the outcomes of the study are more relevant for the analysis of trends in the size and structure of kin networks than in their exact size.

The average number of descending kin (children, grandchildren, nephews and nieces) does not undergo fundamental changes, except for the number of nephews and nieces which, for the population as a whole, increases from 5.0 in the nineteenth century to 7.8 in the 1950s. Despite the recent changes in fertility, the average numbers of children, grandchildren, and nephews and nieces among all age groups in 1990 are still high as or even higher than in the early nineteenth century, mainly as a result of the strong decline in child mortality.

The average numbers of ascending kin (parents, grandparents, uncles and aunts) for the population as a whole changes only little in absolute terms; only the number of uncles and aunts increases considerably. However, fundamental changes by age group take place, with sharp increases especially for the younger age groups.

The average number of lateral kin (siblings and cousins) shows the largest changes in absolute numbers. Again, developments over time differ between age groups: the younger the age group, the earlier the maximum size of this kind of kin network is reached and the earlier it decreases again owing to the lowering of the fertility levels.

It is practically impossible to estimate the total size of the kinship network for the period 1830–1990 by combining all types of relatives discussed in the foregoing, because of the differences in the restrictions made for each type of kin. Nonetheless, the total average number of relatives rises notably from 1830 onwards; between 1930 and 1960 the total number of relatives reaches its highest value. The elderly undergo almost a doubling of the size of their network. Their network of grandchildren, nephews and nieces, and cousins is now relatively large, compared to other age groups and other time periods. Thus demographic changes in Dutch society during the last 160 years have significantly affected the kinship configuration.

Notwithstanding the limitations of the genealogical database, the potentials of it have not yet been completely exploited. The collected data can also be used for historical longitudinal studies, identifying kin experiences of individuals followed over the life-cycle. A subject like the variability of the distribution of numbers of kin might also be discussed on the basis of the data (cf. Le Bras 1982). For the future, one may hope that more accurate genealogical data will become available, but it would be an illusion to believe that this will happen soon. The situation will not change until genealogists become aware of the importance their work might have for our knowledge of demographic history, a development that may perhaps be encouraged by studies showing the use that can be made of existing genealogies (Dupâquier 1993).

NOTES

Wendy Post, Frans van Poppel, Evert van Imhoff and Ellen Kruse, Netherlands Interdisciplinary Demographic Institute (NIDI), The Hague, The Netherlands. This is a revised version of a paper presented at the Twentieth Social Science History Association Meeting, 16–19 November 1995, Chicago. Financial support by the Research Program for Population Studies of the Netherlands Organization for Scientific Research (NWO, grant no. 18.063) is gratefully acknowledged.

1 The table does not include persons born before 1730 (5507) because this study focuses on developments in family structures since 1830, and we assumed that people born prior to 1730 were no longer alive a century later. Also missing in the table are 341 people whose sex was not known, and 641 children born after 1989.
However, records with missing birth-year can still be used to link ego to other kin members.

In spite of this, the genealogies do capture the proper time trend.

REFERENCES


