Available online at www.sciencedirect.com
ScienceDirect
History of the Family 15 (2010) 109–115
The History of the Family

The Antwerp COR*-database: A unique Flemish source for historical-demographic research

Koen Matthijs *, Sarah Moreels

Katholieke Universiteit Leuven, Centre for Sociological Research, Parkstraat 45 / Post box 3601, 3000 Leuven, Belgium

Abstract

This note describes the origin, the history, structure and characteristics of a recently constructed Flemish (the Northern, Dutch-speaking part of Belgium) historical-demographic database. The so-called Antwerp COR*-database offers a unique combination of features: it spans nearly seven decades (1846 to 1920) and consists of information drawn from the population registers and the vital registration records (birth, marriage, and death) of the whole district of Antwerp. Every person whose family name starts with the letter combination COR* is selected in the database. The database covers three linked generations and contains micro-data on the individual level (life courses), intermediary data on family patterns, and macro-data on ecological characteristics.

Keywords: Flemish database; Nineteenth century; Population register; Vital registration records; Intergenerational; Longitudinal; COR*-database

1. Introduction

The international community of social historians, historical-demographers and -sociologists is well aware of the many benefits of the high-quality population registration and vital registration data available in the Low Countries (Belgium and the Netherlands). Studies in Wallonia by Alter (1988) and Oris (1988) have pioneered a successful synergy between historical and social-scientific research questions and methods. In Flanders, the available sources remained very underused until recently. Since the mid-1990s, the Research Group of the Family and Population (Centre for Sociological Research) of the K.U.Leuven (Belgium) has been building up expertise with respect to the collection and analysis of vital registration data (Matthijs, 2001; Van de Putte, 2005) and data from the population registers (Van Bavel, 2002). This project has been funded by K.U. Leuven and Flanders Research Foundation (FWO-Vlaanderen).

2. Research area and research period

The database contains longitudinal and intergenerational data at the individual level for the whole district of Antwerp. From a demographical point of view, Antwerp was the logical first choice because it was the fastest growing Belgian city in the 19th century. By 1900, more than 273,000 inhabitants were living in the port city (1800: 56,000; 1846: 88,000), mainly as a result of massive immigration (Kruithof, 1964; Winter, 2009). This demographic dynamic generates our sociological research questions. The process of migration was indeed heterogeneous: some migrants came from the neighbouring rural areas, others from just over the Belgium-

* Corresponding author.
E-mail address: Koen.Matthijs@soc.kuleuven.be (K. Matthijs).

1081-602X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.hisfam.2010.01.002

Netherlands border, yet others from much further; some came alone, others with their entire family; some returned soon, others stayed their whole lives; some came to work, others to find a partner, yet others with unclear motives. Due to their varying socio-economic and cultural backgrounds, the various groups of migrants wrote a different demographical history with respect to fertility, marriage, divorce, family formation, family networks, chain migration and death.

In the pilot project, the data collection and database construction were limited to the city of Antwerp. The sampling and data collection procedures were tested in order to refine the method before expanding the scope to the surrounding district (see Van Baelen, 2007). As of 2006, the dataset has been enlarged to the Antwerp suburbs (13 municipalities) and afterwards to the whole Antwerp district (another 48 municipalities). This wider geographic coverage has two great advantages: (1) the probability that individuals can be linked increases (which allows to investigate internal migration movements) and (2) the different geographical settings (rural versus urban) in the Antwerp district facilitate comparative research Fig. 1.

Opting for the research period 1846–1920 has a pragmatic basis. The first official census of the independent Kingdom of Belgium took place in 1846, and this year also marks the opening of the first Belgian population register. More importantly however, during the second half of the 19th century, the fertility transition reached cruising speed (Lesthaeghe, 1977). In Belgium, the total fertility index fell from between 4 and 5 around 1850 to less than 2 around 1940. With the Antwerp data, it is possible to reconstruct and interpret the transition, including the initiation, the timing and the intensity of fertility and mortality trends. The processes can be investigated for the native born and foreigners, for internal and external migrants, in an urban or semi-urban environment, and for several generations. Given the fact that the groups of migrants are being followed over a period of time, it is possible to compare successive generations (the first, second and sometimes even the third generation) with each other and with the native population. The intergenerational structure is unique for the Belgian sample. The database enables one to address research questions regarding social and cultural changes which could not be considered with any other Flemish database.

3. Sources

The two main sources for this database are population registers and vital registration records. The Belgian population registers are a high-quality source that enables us to follow individuals, and their family, over time. From 1846 onwards, all Belgian municipalities were obliged to maintain a population register. This register is a repertory of demographic and social information on all official inhabitants of the municipality, and the information is stored per housing unit. For each house, all changes were recorded, such as births, deaths, marriages, divorces but also internal moves and external migration (Gutmann & van de Walle, 1978; Leboutte & Obotela, 1988). Vital registration records refer to birth, mortality and marriage events. These certificates contain a number of relevant social and cultural characteristics of the people involved (the baby, the deceased, and the migrant), of the direct environment (e.g. the parents) and the witnesses. The registration of births, marriages and deaths began in Belgium on 17 June 1796. Apart from a limited number of adjustments, little changed in the content of the marriage certificates in the 19th century. Population registers and vital registration records are usually well preserved in Belgium and are stored in state or town archives, and for smaller municipalities, in municipal archives.

4. The Antwerp COR*-database

4.1. The letter sample

Collecting all information from the population registers and vital registration records for the whole district of Antwerp during the period 1846–1920 was not feasible. Thus, after an extensive evaluation of alternative data gathering strategies, we opted for a letter sample. A letter sample – also used in historical projects in France (Dupâquier & Kessler, 1992; Bourdelais, 2004), Canada (Ornstein & Darroch, 1978; Darroch, 2003) and Wallonia (Alter, Neven & Oris, 2002) – includes all information about persons whose family name starts with a given combination of letters. Three reasons were decisive for choosing a letter sample. First, by making use of a letter sample, the data collection is simplified, thus leading to enhanced reliability. By giving clear and simple instructions to the people collecting the data, mistakes are minimized. Second, a letter sample makes it easier to deal with imperfect data sources and missing data. The reconstruction of socio-demographic life courses involves linking information that is scattered in many different time periods, places, and sources. When a piece of the puzzle is missing, it is difficult to link information about earlier parts of the life course to later parts. Especially migration is a challenge.

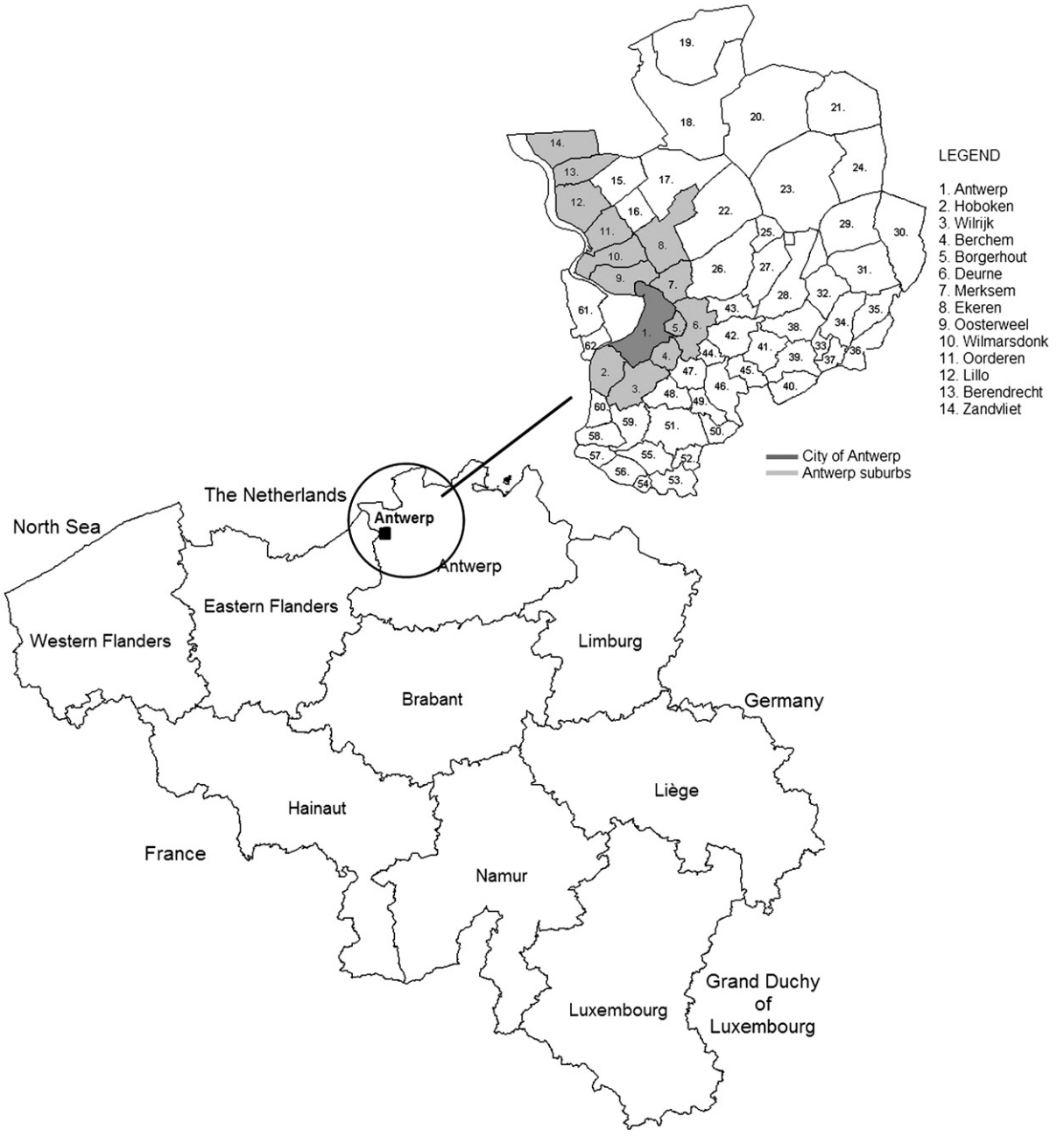

Fig. 1. Map of Belgium with the district of Antwerp, 1900.

With a letter sample, it is less difficult to find migrants in different localities, even if we do not have explicit information about migration timing and destination. This information is completely lacking in localities where the population registers got lost. Without a letter sample, it would have been like looking for a needle in a haystack to find out where an emigrant had gone. The bonus of the letter sample is that we are able to fully incorporate migration in our analyses, both as a covariate and as the dependent variable. This is crucial to address the specific aims of our research. Third, our sampling method facilitates the study of the effects of kinship and family networks on demographic behavior.

The letter combination that is the basis of the sample satisfies a number of conditions. The main condition involves geographical distribution: the letter combination has to be well spread over the Flemish territory. The K.U.Leuven Department of Linguistics (Faculty of Arts)

helped us select a letter combination that met this criterion. COR*-names were selected as the best option, because the geographical distribution of COR*-names is very similar to the distribution of the total Flemish population. We also found that the distribution of people with COR*-names is representative (for 19th-century Flemish population) of a number of socio-demographic characteristics and is language sensitive as well (which is especially important with regard to foreign peoples). Moreover, legibility problems (c/t confusions or c/k transformations) are negligible (Van Baelen, 2007). The sample size is 0.4% of the Flemish population, which means that small geographical areas cannot be investigated separately with this dataset.

4.2. The process of data collection

The data collection began with the population registers and has been supplemented with information from the vital registration records (birth, marriage and death certificates). Using the population register, it is possible to gather more information in less time. Also these registers contain information about migration, which is lacking completely in the vital registration records. Information about migration is essential in order to estimate the population at risk for longitudinal analyses and in order to calculate accurate demographic rates. We also complemented the information from the population register with data from the vital registers of birth and death, because these certificates are more reliable (Gutmann & Van de Walle, 1978) and in some cases, provide additional data. Stillbirth, for example, is not registered at all in the population registers. From the vital registration records of birth and death, we checked if births or deaths were missing in the population register. From the register of marriage, we checked systematically the marriage certificates of all COR*-people. The marriage certificate establishes the marriage in law. In addition to information on the date and place of the solemnisation of the marriage, it contains a variety of socio-demographic data about the bride and groom themselves, their parents and the witnesses.

The left column of a Belgian population register always contains the family names of people who lived at that address during the years that a population register was open. The first population registers started in 1846 and were followed by new registers in the years 1856, 1866, 1880, 1890, 1900 and 1910. We selected a 'house' (a house may consist of several households) if a COR*-person had been living there. All COR*-persons and their kin, present in that household, were entered in Microsoft Access. Because not all population registers are yet microfilmed in the district of Antwerp, we opted for a double track in data collection. The registers that are available on microfilm were consulted in the state or city archives of Antwerp. The other population registers (mostly from 1866 onwards for the smaller municipalities) were consulted in their original format in the register's office. We took digital photos of the selected pages.

Tables 1 and 2 present the number of records in the Antwerp dataset after the data collection: 14,537 houses which contained 56,386 observations of persons and 126,403 events, and 14,697 COR*-certificates were entered. The total number of unique family names amounts to 8032, of which 675 were COR*-family names. The number of two-generation families (parents and the children) totals 2650 (referring to 9486 children), for three-generation families (grandparents, parents and children) the corresponding figure is 1157 (involving 4381 children).

4.3. Database construction

After the data collection, all observations for a given individual must be linked and ordered chronologically

Table 1
Data entry of population registers, district of Antwerp, 1846–1920.

	Houses	Observations	Events [a]
City of Antwerp	9868	33,023	93,198
All suburbs of Antwerp [b]	2891	13,836	20,921
District of Antwerp [c]	1778	9527	12,284
Total	14,537	56,386	126,403

[a] Events are immigration, emigration, moves in the same municipality, death, marriage or widowhood.
[b] All suburbs of Antwerp, except the city.
[c] The total district of Antwerp, except the city and the suburbs.

Table 2
Data entry of vital registration records, district of Antwerp.

	Birth certificates (1821–1906)	Marriage certificates (1806–1913) [a]	Death certificates (1836–1906)
City of Antwerp	3448	1224	3109
All suburbs of Antwerp [b]	1547	483	1091
District of Antwerp [c]	1987	411	1397
Total	6982	2118	5597

[a] For the period 1907–1913 the marriage certificates are only collected for the city and the Antwerp suburbs; for the other municipalities of the district, information was gathered until 1906.
[b] All suburbs of Antwerp, except the city.
[c] The total district of Antwerp, except the city and the suburbs.

into a life course. This is a complex and time-consuming process, but the result of this process – individual life courses – is a treasure for historical demographic researchers. Before starting this nominal record linkage, it was necessary to clean and standardize the collected data. First names, last names, places of birth and dates of birth are used to link the nominal records into individual life courses. We not only cleaned the first names, but we also standardized them. This was necessary because there was little consistency in the spelling of first names. Jozef Adolf August (Flemish variant) was also noted as Josephe Adolphe Auguste (French variant). After testing different strategies, we opted for standardizing the names by converting them to the Latin variant (for example: Jan and Jean become Joannes, Frans and François become Franciscus). Last names were standardized according to 23 simple rules (Van Bavel, 2002). These rules take common spelling and pronunciation variations into account. Examples are: 'ca' becomes 'ka', 'ce' becomes 'se', 'ck' becomes 'k' and 'ey' becomes 'ij'. The place of birth is converted to a five digit standard municipality code of the National Institute of Statistics (Foulon, Poulain, Van Goethem & Lux, 1981; Nationaal Instituut voor de Statistiek, 1983). For foreign municipalities we constructed a new code. All foreign codes start with a unique country code of four digits, followed by a unique place code, again of four digits. In this way, we can see if a municipality is Belgian or foreign. For the place codes of England/Wales and the Netherlands, we used the standard codes of the two countries.

The linkage of different observations of the same individual was done in three steps. In the first step we selected potential pairs of observations to be linked. Next, we evaluated in detail these potential linkages. Finally, we assigned new common identification numbers to the linked records. To select the potentially linked records, we made 13 Access-queries. In these queries, we compared two different records of persons with each other. When there is enough resemblance, they were selected as a potential link. In the second step, the potential linkages were automatically and manually evaluated. Whether or not to link, was first decided automatically. Both positive (linking) and negative (not linking) decisions were made by the computer algorithm. Previous research (Van Bavel, 2002) already showed that it is possible to automatically make negative decisions about many potential links in case of very low match scores. After the automatic linkage, the remaining potential links were evaluated manually. By making use of a coupling form in Access, which presents all information on the possible linked records together on the screen, the researcher decided via a command button whether a certain couple is a link or not. In the third step, new identification numbers are automatically assigned to the records so that each unique identification number refers to one individual life course. Around thirty thousand complete life courses were reconstructed out of the 56,386 observations, amounting to more than 800,000 person-years.

Inevitably, when collecting and entering that much data, mistakes are made. We did a reliability check with the registers of birth and death. Moreover, the vital registration records provide additional information, which is not included in the population registers. By linking the vital registration records with the population registers (through cleaning, selecting potential linkages, automated and manual evaluation for every type of certificate), it is possible to expand the number of relevant demographic and sociological research questions substantially. Vital registration records indeed contain data which is not available in the population register. From birth certificates, we obtain information on the date and place of notification, the recognition and legitimation of illegitimate children, the hour of birth, information on the two witnesses (first name, last name, occupation, age, domicile and kin relation to the newborn child), information on the informer (mostly the father), illiteracy and additional information on occupation. The death certificates give additional information on the date and place of notification, stillborn children, the hour of death, detailed information on the age of the deceased (in years, months and days), information about the two informers (first name, last name, occupation, age, domicile and kinship with deceased), illiteracy and additional information on occupation. The marriage certificates include information on possible divorce, the occupation of the parents of the non-COR*-person (important for research on social mobility), date and place of the wedding, information on the witnesses of the marriage, marriage settlement or not, the legitimation of children and information on illiteracy that is not included in the population registers. After the linkage of records for the same individuals, we still need to link family members. Identifying the family members was also automated after carefully analyzing all kinship relations (Van Baelen & Matthijs, 2007).

After the linkage and the vital registration check, the data-entry file must be converted into a database that is structured to facilitate analyses. The most important difference between the file and the database is that the data-entry file was constructed in order to resemble the source, whereas the database groups all linked data per person and per type of characteristic. Each person has

fixed characteristics which do not change in the life course: last name, first name(s), date of birth and death, and place of birth and death. Other characteristics are variable: occupation, civil status and nationality. These characteristics require flexible formats. Therefore we designed a structure of five connectable subsets of the data: one table contains all fixed characteristics and the ID (identification), the others are comprised of variable characteristics. There are five tables: the table with the fixed characteristics, the sequences table, the marriage table, the kinship table and the events table. Events are: immigration, emigration, death, marriage, move (to another house in the same village or city), coming in (from another house in the same village or city), widowhood, birth, occupation and nationality, and start and end of population register.

4.4. Contexts for comparison

The database is also expanded with contextual data. Aggregate data from the agricultural (1846), the industrial (1896 and 1910), the trade (1910) or the population censuses provide additional information on the economic and socio-demographic situation of the district of Antwerp during the 19th century. Moreover, a quarter table which indicates changing quarter boundaries through time is also available for the city of Antwerp. These additional socio-economic and cultural information at the regional and local level offer several interesting contexts for comparison.

(1) Since the data relate to the entire district of Antwerp, urban processes can be compared with semi-urban and rural processes: (a) native versus foreign-born populations, (b) metropolitan versus suburban long-term developments, (c) intra- and intergenerational comparisons, and (d) all of which can be situated within both national and international contexts;
(2) The Antwerp immigration was often motivated by the employment opportunities at the port. Dock work is physically demanding and consequently there was a relatively high male labour migration to Antwerp. That as opposed to, for example, the developments in Ghent and Aalst where there was a greater 'import' of female labour in the textile sector. That difference allows one to make relevant comparisons of the interaction between gender, work sector and migration status. This is possible because the Leuven research group also has historic–demographic information on a number of other Flemish cities, mainly from the vital registration records;
(3) For the period 1846–1920, the study group has access to all the information contained in the population register of Leuven. Therefore, it is also possible to draw a comparison between Antwerp and Leuven. Our comparative analysis may also incorporate the Walloon city of Verviers (data available from George Alter, Alter, 1988). These three cities provide a significant range of cultural, social and economic characteristics;
(4) Moreover we are able to bring an international dimension to our comparative research. The Historical Sample of the Netherlands (HSN)-project for instance shares a number of features with the Antwerp sample (Kok, Mandemakers & Bras, 2009);
(5) A crucial sociological question is whether modern integration today is following (or will follow) the same path as in the past, or whether modern processes are unique. The latter 'uniqueness school' is based on the fact that the structural circumstances are now different (the welfare state is fully developed, there is a social security system, the migration policy is different). One of the ways of contributing this issue is to carry out a thorough study of the historical-demographical behaviour with the help of micro data.

5. Future agenda

The unique features of the Antwerp COR*-database are: it spans nearly seven decades of time (1846 to 1920) and consists of information drawn from the population registers and the vital registration records of the whole district of Antwerp (Belgium, Flanders). The Antwerp database covers three generations and the members of these generations are coupled. The database contains extensive micro-data on individual life courses and family patterns, and this information on individuals and their family is connected to a rich array of contextual data. The construction of this complex dataset was expensive and time-consuming, but the potential pay-offs are significant in terms of a range of sociological issues that may be addressed and the ability to apply powerful statistical techniques to the analysis of these issues. The database will facilitate the investigation of social and cultural changes in greater depth than could be accomplished with any other Flemish database. Using this database, life events (fertility, migration, and mortality), sibling relations, migration flows and family systems can be studied on an intra- as well as on a intergenerational basis. In sum, the added value for new historical-demographical research using the Antwerp

COR*-sample, is (1) the linkage of a large number of data elements (from different sources) relating to individuals, (2) the reconstruction of family characteristics, sibling systems and life courses (migrant versus non-migrant, urban versus semi-urban and rural), (3) the augmentation of individual information with contextual data, and (4) the development of an intra- and intergenerational longitudinal perspective (period 1846–1920).

In order to fully expand the potentials of the Antwerp COR*-sample, we plan to extend the sample to other regions, first to adjacent, later to remote regions. In that manner, causes and consequences of internal and international migration patterns and processes can be studied more accurately. Enlarging the time-span and the geographic scale makes it easier to distinguish context-specific from universal mechanisms behind behavioral outcomes and social change (for a theoretical framework, see Matthijs, Kok, Engelen, van Poppel & Mandemakers, 2006).

In the long run, we aim to integrate the Antwerp COR*-sample with the *Intermediate Data Structure* (IDS) protocol (Alter, Mandemakers & Gutmann, 2009). To successfully analyze demographic behavior across different contexts (periods, localities, and population subgroups) datasets have to be made comparable. Separate databases need to be converted into a common structure, clearly separating individuals (and their relations), events (e.g. marriage, death, and migration), and contexts (addresses, localities, and regions). This is a major objective of the IDS-network.

Those who wish to learn more on the COR*-database, to collaborate on a joint research project, to analyze historical-demographical themes in a comparative way, and to publish on these issues in international journals, are kindly invited to contact Koen Matthijs (Koen.Matthijs@soc.kuleuven.be). Experience teaches that the international comparison of life events, life courses, family systems, migration patterns and related issues, opens up a fascinating array of insights into the past.

References

Alter, G. (1988). Family and the female life course. The women of Verviers, Belgium, 1849–1880. Madison: The University of Wisconsin Press.

Alter, G., Mandemakers, K., & Gutmann, M. (2009). Defining and distributing longitudinal historical data in a general way through an intermediate structure. *Historical Social Research*, *34*(3), 78−114.

Alter, G., Neven, M., & Oris, M. (2002). *Stature in Transition. A Micro-level Study from Nineteenth Century Belgium. Working Paper Series' Early Life Conditions, Social Mobility and Longevity*, *6*.

Bourdelais, P. (2004). L'enquête des 3000 familles: un premier bilan. *Annales de Démographie Historique*, *1*, 5−6.

Darroch, G. (2003). Revisiting the Old "New Social History": Social closure, migration and social mobility in South-Central Ontario. A regional study through linked "letter" samples, 1861–1871. *IMAG Workshop 10–11 November 2003, Longitudinal and Cross-Sectional Historical Data: Intersections and Opportunities*. Montréal, Québec: International Microdata Access Group.

Dupâquier, J., & Kessler, D. (1992). L'enquête des 3000 familles. In J. Dupâquier, & D. Kessler (Eds.), *La société française au XIXe siècle. Tradition, transition, transformations* (pp. 23−61). Paris: Fayard.

Foulon, M., Poulain, M., Van Goethem, B., & Lux, B. (1981). *Répertoire historique des communes Belges De 1831 à nos jours / Geschiedkundig overzicht van de Belgische gemeenten van 1831 tot heden*. Louvain-la-Neuve: Université Catholique de Louvain, Département de Démographie/cabay.

Gutmann, M., & van de Walle, E. (1978). New sources for social and demographic history: The Belgian population registers. *Social Science History*, *2*(2), 121−143.

Kok, J., Mandemakers, K., & Bras, H. (2009). Van geboortebank tot *collaboratory*. Een reflectie op twintig jaar dataverzameling en onderzoek met de HSN. *Tijdschrift voor Sociale en Economische Geschiedenis*, *6*(4), 3−36.

Kruithof, J. (1964). De demografische ontwikkeling in de XIXde eeuw. *Genootschap voor Antwerpse Geschiedenis, Bouwstoffen voor de geschiedenis van Antwerpen in de XIXde eeuw* (pp. 508−543). Antwerpen: Anversois.

Leboutte, R., & Obotela, R. (1988). Les registres de population en Belgique. Genèse d'une technique administrative et d'une source de démographie historique. *Handelingen van de Koninklijke Commissie voor Geschiedenis*, *3–4*, 285−305.

Lesthaeghe, R. (1977). *The decline of Belgian fertility.* Princeton: Princeton University Press.

Matthijs, K. (2001). *De mateloze negentiende eeuw. Bevolking, huwelijk, gezin en sociale verandering*. Leuven: Universitaire Pers Leuven.

Matthijs, K., Kok, J., Engelen, T., van Poppel F. & Mandemakers, K. (2006). A turn of fate. Life courses in the Low Countries between 1850 and 1940. *WOG Working paper* 2006. http://soc.kuleuven.be/ceso/historischedemografie/resources/pdf/WOG%20working%20paper01.pdf.

Nationaal Instituut voor de Statistiek. (1983). *Geschiedkundig register van de Belgische gemeenten van 1831 tot heden. Alfabetische rangschikking: benaming van de gemeente*. Brussel: Nationaal Instituut voor de Statistiek.

Oris, M. (1988). Fécondité légitime et illégitime: les indigents de Huy (Belgique) de 1815 à 1875. *Annales de Démographie Historique*, 141−155.

Ornstein, M., & Darroch, A. (1978). National mobility studies in past time: A sampling strategy. *Historical Methods*, *11*(4), 152−161.

Van Baelen, H. (2007). *Constructie van een historisch–demografisch longitudinale database. Methodologie van de Demographica Flandria Selecta.* Leuven: Centrum voor Sociologisch Onderzoek.

Van Baelen, H., & Matthijs, K. (2007). Comparing two methods to reconstruct the kin network outside the household with 19th century Antwerp data. *Journal of Family History*, *32*(2), 149−160.

Van Bavel, J. (2002). *Van natuurlijke naar gecontroleerde vruchtbaarheid? Geboortebeperking in Leuven, 1846–1910.* Leuven: Universitaire Pers Leuven.

Van de Putte, B. (2005). *Partnerkeuze in de 19de eeuw. Klasse, geografische afkomst, romantiek en de vorming van sociale groepen op de huwelijksmarkt*. Leuven: Universitaire Pers Leuven.

Winter, A. (2009). *Migrants and Urban Change: Newcomers to Antwerp, 1760–1860.* London: Pickering & Chatto.