

**INTRODUCTION:
DEMOGRAPHY AND DATABASES
IN THE EAST AND THE WEST**

Gunnar Thorvaldsen,

Elena Glavatskaya,

Mikolaj Szoltysek

A conference held on the border between Asia and Europe inspires dual vision: looking eastwards as well as westwards. Our main aims with this conference were comparing source materials, databases and research results, as well as creating new opportunities for collaboration.

In terms of historical sources, a comparison of the quite accurate Chinese Han dynasty census from 2 CE with the heterogeneous descriptions of property in the British Domesday Book a millennium later, makes it clear that the history of population lists is longer in the East than in the West. For military and taxation purposes, the Mongols extended the Chinese enumerations to Russian territory in the thirteenth century. Russia continued this tradition, and the first census-like (tax) lists seen by West Europeans after the end of the Roman Empire were probably Russian in origin. The German ambassador to Moscow, Baron von Herberstein, in the sixteenth century reported that censuses were taken biannually. In the 19th century, bilateral contacts between statisticians and international statistical conferences spread the practice of nominative census taking first to Great Britain in 1841, next to the United States in 1850 and later even further afield. This introduction briefly browses the history of the nominative source material, and how many of these original census manuscripts have been preserved, digitized, transcribed and disseminated as historical microdata – oftentimes through international cooperation (Thorvaldsen 2016). Interspersed are references to the contents of the other articles in this volume.

Highlights from the history of the sources

Unfortunately, nearly all census manuscripts from the first Russian imperial census of 1897 and the Soviet censuses from 1920 onwards were destroyed in order to save archival space. The authorities even destroyed the 1937 census *aggregates* because the results were politically unacceptable. The preserved manuscripts are a few local 1897 questionnaires; the computerized versions for the most recent decades; and significant parts of the 1926–27 Polar census – the most comprehensive census taken anywhere, with details about the economy and qualitative descriptions of the northern indigenous peoples (Anderson, 2011; Glavatskaya, Kljukina-Borovik, 2013). The archival situation is better for the comprehensive collections of Russian tax lists and parish records. Central State Archives in Moscow and St. Petersburg as well as most regional archives hold *revizkie skazkie* (tax revisions) from the 1720s to 1858, some of them with census-like individual level data about entire households. Covering the same period, as well as the later period from the 1850s until the revolution in 1917, the archives contain parish registers that give vital information about the baptized, the married, the buried and the divorced for most parts of the Russian Empire. Most localities are represented, and there are even records for religious minorities. This practice stopped after the Revolution, but paradoxically the new atheist rulers preserved the parish registers, while destroying the secular censuses. Few of the ministerial records in Eastern Europe and beyond have been transcribed for computer analysis, but fortunately new samples are now being brought forward in the Urals, Altai and Transylvania.

Some countries in-between the East and the West have archived historic population registers and/or a combination of censuses and church registers. Denmark, Iceland and Norway which used to make up the Danish Kingdom, demonstrate the common history of censuses and church registers and their employment for demographic analyses, with the full count 1801 nominative census as a highlight (Thorvaldsen, 2017). The centuries long series of longitudinal source material enabling the construction of detailed population registers from catechismal and other church records makes the situation different and superior in Sweden and Finland.

Efforts to create an overview of the size of the population at a given time and demographic developments occurred in parallel in different countries. This can be explained both ideologically and with practical

needs. Theological opposition traditionally blocked work on population statistics - man should not control God. The basis for such scepticism in the scripture can be questioned, even though there are biblical passages warning about numbering people for military purposes, see for instance Samuel II, chapter 24. Perhaps practical reasons were more important - the priests knew who would have to carry out the work. The bishops' visitation reports show how they had to press the priests to make them keep the church books properly (Dyrvik, 1983). A further reason for resistance was political, as the proper construction of central statistics requires a strong central state power capable of requiring the necessary demographic data from local and regional leaders. The latter felt their position threatened by transferring control over their subjects to the central level, cf the gentry's opposition towards holding censuses in the UK in the 18th century.

The ideological element of the early social sciences and population statistics is also important. The Cartesian research method recommends studying each component of the object under study carefully, just like the statistical aggregates were based on the priests' information about each individual. Descartes prescribed how to compile anatomical details into an overall presentation of the research object, in analogue with the way vital statistics provide aggregate overviews of population developments. Thus, the methodological recommendations were accompanied by biological analogues about how the body parts relate to the whole man like the individual relates to society (Olson, 1993).

Practical concerns indeed triggered the collection of census data and vital statistics. Military needs obviously motivated the taking of *male* censuses. Data collection was part of the state's intention to overview human resources prescribed by 18th century leading ideology, mercantilism. The practical goal was to promote business development in each country, for instance to increase revenues for military activities. Population growth became an instrument because the people were seen as an input factor in the economy as well as a criterion to measure whether the mercantilist policy was successful. Towards the end of the 18th century, officials would subordinate the economic objectives to the political objective of promoting the well-being of the population. "The aim was to form administrators or bureaucrats who should be concerned with the welfare aspect ... including the moral and physical health of the population." (Olson, 1993, p. 134). The nominative, full-count 1801 census and the ordinance of 1812 to the priests about church books with fixed headings were instruments for such reforms in the Danish Kingdom (Dyrvik, 1983).

Before the 20th century, Russian administrators had experience with aggregating census data primarily from the 1897 census, which covered the whole empire except Finland, and from censuses run in many cities in 1870s and 1880s. However, on the individual level, they had for nearly two centuries worked on the revizkie skazkie (census like tax revisions), which had been introduced by Peter the Great in 1718 – the 10th and last was organized in 1858. In seven of them, the whole population was in principle included, making it possible to study the household and family structure, as Maria Markova does based on the fifth revision for urban sites in and around St Petersburg in the present volume. The relationship variable in the revisions make it possible to see what persons in a household belonged to the same family and on this basis to classify some households as multi family. The revision also provides data on age, occupation and marital status of the persons listed, including women. Also, the historical demography group at Ural Federal University transcribe and research the Revizkie Skazki from 1852 and 1858 about the Obdorsk¹ Samoeds (contemporary Nenets) and Ostiaks (contemporary Khanty). People were obliged to pay taxes for the men listed until the next revision including their serfs, thus also for those who had passed away in the meantime. This was the basis for Gogol's masterpiece novel *Dead Souls*. The plan of Chichikov, Gogol's hero-villain, was to buy up the "dead souls" at reduced rates, acquiring a list of fictitious but legally valid serfs. These he meant to use as security when investing in an estate and alive serfs. We can find these lists in archives all over the previous Russian Empire and scanned images are available on the Internet, for instance from Tobol'sk collection of the Tjumen' State Archive.

A number of source-critical issues can be raised, both about general and detailed aspects of the church books, the population censuses and other demographic source materials. For instance, when comparing place and date of birth of individuals in the church book with census entries, they often do not match (Thorvaldsen, 1996a, p. 63-64). However, divergences can also be due to the fact that people used a church in the neighbouring parish because the distance there was shorter, or the parish borders were adjusted. The most serious problem may be that the oldest church books lack many entries. This applies mostly to the funeral lists, and to the least extent in the list of marriages. The church

¹ The 19th century Obdorsk region included the Yamal Peninsula together with the Eastern slope of the Polar Urals, part of the Gyda peninsula and the gulfs of the Ob' and Taz Rivers with its tributaries.

books recorded the ecclesiastical acts, not primarily aiming to overview population developments. For example in Norway, the roughness of the oldest church books is revealed by formulations such as “Earthed 5 Bodies” or “Buried a Fin wife “. The under-enumeration of the dead is partly due to the fact that people were most keen to pay for a church burial when important persons died (Høgset, 1990). The vocabulary in the church book is usually less precise than in the census, often indicating a rough status category like “cottar” rather than a more specific occupational function such as “fisherman”. Researchers have compensated for this by linking church book and population record data at the individual level (Engelsen, 1983). Cause of death is another type of imprecise or often missing information. Methods to study migration in spite of the many missing migration records in the Norwegian church books are presented in the article by Gunnar Thorvaldsen. Georg Fertig analyses how the rich collections of demographic data based on vital events records and address registers existing throughout Germany can be used in demographic research. Marius Eppel and Oana Sorescu-Iudean present a more specific picture of nominative sources for the study of the Transylvanian clergy.

Historical databases

In addition to church records, archives spread across the European continent contain large series of censuses and census-like sources. Many of these have been transcribed digitally and included in the comprehensive Mosaic database (Szołtysek and Gruber, 2016). In the present volume, Siegfried Gruber bases his study of marital fertility in Albania on one Mosaic component, the census taken there during the Austro-Hungarian occupation in 1918. Some countries in-between the East and the West have archived historic population registers in part based on the church registers. Sweden, Finland, Belgium and the Netherlands have paper-based population registers from the eighteenth or nineteenth century onwards. Three Swedish databases contain longitudinal population registers for five Swedish regions, including the capital city of Stockholm between 1876 and 1928. These databases offer details on migration, educational levels, etc. that are seldom found in historical records. Their richness makes digitization resource demanding – with traditional transcription methodology they may not cover Sweden during this or the next century. An impressive bibliography of recent publications based on the linked church records from Swedish regions,

covering individuals from 1680 to 1950, can be found at their website <https://www.umu.se/en/centre-for-demographic-and-ageing-research/>. With respect to transcription efficiency, the digital techniques developed in cooperation between historical demographers and computer scientists at the Autonomous University in Barcelona, which Joana Pujades Mora and her team present in their article, are most promising. To compensate for regional selectivity in the meantime, the Swedish censuses from 1860 to 1930 are being transcribed, encoded, linked and made available with national coverage. The Historical Sample of the Netherlands contains longitudinal data following a representative group of 78,000 individuals during the period 1812 to 1922. Peter Öri links household lists and vital events in order to build a historical population register, a preliminary version of which he uses here to study marriage patterns in the Hungarian village of Zsámbék from the 18th century to World War II. And Mark Gortfelder finds a partial effect of child mortality on fertility (on stopping, but not spacing of births) by employing the Family Registry of the Estonian Republic 1926–1949, which also contains retrospective data on persons born before 1926. Hilde Sommerseth and Evelien Walhout present a wider scope of the mortality issue in their study of death causes in the mid-Norwegian town Trondheim, which is also an important contribution to the SHIP project.

Elena Glavatskaya and her team members Dmitrii Bakharev, Alexander Bobitskiy, Elizaveta Zabolotnykh and Anastasia Vishnevskaya analyse the main marriage patterns and infant mortality in two articles about late 19th – early 20th century Ekaterinburg, for the latter theme also the rural surroundings. They used aggregate census data along with microdata extracted from several city parish registers (*metricheskie knigi*), transcribed into the database “Ural Population Project”. Both the majority Russian Orthodox Church parishes and several religious minorities: Catholics, Lutherans, Muslims and Jews are represented. In all, the database contains about 25000 baptism records, 15000 marriage records and 31000 burial records. The analyses show that marriage patterns depend on religious affiliation as well as social status and migration. They also found that while in the city the high infant mortality rate among the Orthodox majority had minimal changes, in Ural villages IMR decreased as a result from the efforts of the *zemstvo* – a self-government, elected, sub-provincial institution introduced in 1865.

The basis for keeping vital statistics was reformed with the introduction of civil registers, for instance in Russia after the 1917 Revolution and in Norway from 1905. The inappropriate use of such instru-

ments can be seen there during World War II when the Germans in February 1942 attempted to establish a central population register (Søbye, 1998). A longitudinal population register combines cross-sectional data with vital register data of the kind found in the church books. It is usually based on a census as a starting point and is updated with demographic events and migration. The primary aim of the Germans was to control passenger traffic in all municipalities, but the register could also be used to assemble statistics about births, marriages and deaths, something Statistics Norway stressed in its comments on the reform, which they were then able to delay until after the War by blaming quality problems. In connection with the census of 1950, all municipalities updated the population records with information from the population census manuscripts. This census is presently the first full-count, closed historical census in the world to be transcribed (Thorvaldsen, 2018).

If the population register contains all the variables that underpin vital statistics and aggregates from population censuses, it will significantly simplify the collection of basic data. The statistical agency can stop organizing questionnaire-based censuses and instead extract data from the population register by combining it with additional databases. These can be linked to the central population register at the individual level with a unique identity number as the linkage key. This is the reason why Denmark held its last form-based census in 1970. In addition to the registry method saving costs, its main advantage is that with little extra resources, the computer can construct statistical overviews of the population's composition and the ongoing changes as often as desired. However, it would be desirable to check the registers by sending forms to at least a representative sample of the population. This also provides additional information about potentially missing variables such as housing, commuting and education. Education acquired abroad may be a variable where we will always need form-based updates.

Eighteen European countries used register data at least partly for their enumeration during the 2011 round of census taking, sometimes in combination with data from questionnaires. The purpose was to create census aggregates, although they also constructed nominative lists. In the future, researchers can access both aggregates and (anonymized) census-like micro-data, albeit extracted from registers. The advent of population registers signals the end of the difference between nations focusing on censuses and those focusing on vital register data. The methodology will bring forth new source materials that are more suited to follow over time population segments that are increasingly ge-

ographically mobile, both nationally and internationally, and will produce censuses and vital register data more economically. The downside may be that the population loses the control they had over the information, which they used to provide through the questionnaires.

The need for big data also in historical demographic research inspires the employment of the new information technology methods to rationalize data entry and the combination of sources into longitudinal databases with record linkage. The amounts of available source material are enormous, and even if much has been computerized, even greater collections of census manuscripts, ministerial registers etc are awaiting scanning, transcription and linkage. Just think of the quantities of *metricheskie knigi* (church books) from the vast Russian Empire during the two centuries before the Revolution. Only a fraction of this has been opened up with modern IT so far, and there are small prospects to get resources to transcribe much of this with traditional methods. Therefore, the developments described in Joana Pujades' article with crowdsourcing, word spotting and the use of artificial intelligence - machine learning specifically – are most promising as methods to rationalize digitisation decisively. Fortunately, there are now similar methodological developments at several research centres around the world, and they are also being discussed at the international conferences for historical demography and social history.

References

- Anderson, D., Ed. (2011). *The 1926/27 Soviet Polar Census Expeditions*. Oxford, New York, Berghahn.
- Dyrvik, S. (1983). *Historisk demografi. Ei innføring i metodane*. Bergen.
- Engelsen, R. (1983). "Mortalitetsdebatten og sosiale skilnader i mortalitet." *Historisk tidsskrift* 62(2): 161–202.
- Glavatskaya E., Kljukina-Borovik Ju. (2013). *Ural'skaya ekspeditsiya na Obdorskome Severe: Pripolyarnaya perepis' 1926–1927 gg.* [The "Ural" Polar Census Expedition to the Yamal Region 1926–1927]. Ekaterinburg : Ural Univerity press.
- Høgset, R. (1990). "Begravelsesskikker og trosforestillinger i det gamle bondesamfunnet. - en feilkilde når en bruker de gamle kirkebøkene til å studere dødelighet?" *Historisk tidsskrift* (2).
- Olson, R. (1993). *The Emergence of the Social Sciences 1642–1792*. New York etc, Twayne Publishers.
- Szoltysek, M. and S. Gruber (2016). "Mosaic: recovering surviving census records and reconstructing the familial history of Europe." *The History of the Family* 21(1): 38–60.

Søbye, E. (1998). "Jødeforfølgelsene under den annen verdenskrig. Et mørkt kapittel i statistikkens historie?" *Samfunnsspeilet* (4): 2–12.

Thorvaldsen, G. (1996a). *Håndbok i registrering og bruk av historiske persondata* [Handbook for transcribing and using historic microdata]. [Oslo], Tano Aschehoug.

Thorvaldsen, G. (2016). The East, the West and the future of demographic data. *The future of historical demography : Upside down and inside out*. K. Matthijs, S. Hin, J. Kok and H. Matsuo. den Haag, Acco Leuven.

Thorvaldsen, G. Ed. (2017). *Three centuries of northern population censuses*, Routledge.

Thorvaldsen, G. (2018). Avtomatizaciya transkribovaniya istoricheskikh istochnikov: opyt raboty s materialami perepisi naseleniya Norvegii 1950 goda. [Automating Historical Source Transcription with Record Linkage Techniques. Work in progress on the 1950 census for Norway]. *Istoricheskaia informatika* (1): 94–103.

Thorvaldsen, G. (1995). Longitudinal Sources and Longitudinal Methods – Studying Migration at the Stockholm Historical Database. *Swedish Urban Demography*. Tedebrand and Brändström (eds.). Umeå, Umeå University: 219–250.