SURVEY OF HISTORICAL DATABASES

WITH LONGITUDINAL MICRO-DATA

The second questionnaire

For more information about this questionnaire or questions about entering specific Information, please contact Kees Mandemakers (kma@iisg.nl) and/or Tatiana Moisseenko (tatiana.moisseenko@iisg.nl)

Table of Contents:

The questionnaire comprises three sections:
Section A includes the questions related to the most general and important information identifying the content, scope and provenance of the databases and the information about their creators.
Section B contains more specific and detailed questions about databases, such as the period(s) of observation, sampling design and procedures, data collection, linkage process and others.
Section C contains detailed questions about sources used for the databases: their type, scope, content, state of preservation, etc.

Section A

I. General (identifying) information about the database

1. Title of the database	SCANIAN ECONOMIC DEMOGRAPHIC DATABASE
1.a. Subtitle, which brings meaning to the title (scope, place, time period):	
2. Abbreviation	SEDD

3. Links to website(s):	
3.a. Homepage	http://www.ed.lu.se/databases/sedd
3.b. Get to data	

4. Abstract: describes content of the database. Max. length: 300 words Please indicate: ◦ Scope and main goal ◦ Time and territory covered by data ◦ Sample strategy ◦ Main sources	The Scanian Economic Demographic Database (SEDD) consists of an event database, in which all basic source material is registered, and an applied research database. The event database is based on family reconstitutions for nine rural parishes (Ekeby, Frillestad, Halmstad, Hässlunda, Hög, Kågeröd, Kävlinge, Sireköpinge and Stenestad) and one city (Helsingborg) in Scania, the southernmost county of Sweden, in which information in church records on births, deaths and marriages are linked together to form families. The event database is accessible through the Regional Archives in Lund.

	The research database covers information on the total population in five rural parishes c. 1646 to 2011: Halmstad, Hög, Kågeröd, Kävlinge and Sireköpinge. One of the rural parishes (Kävlinge) was transformed to a small industrial town in the 1880s. It also encompasses one urban city: Landskrona. The description that follows primarily relates to the research database. At its present stage, the research database includes data for 150,000 individuals up to 1968, to which data from central registers are linked for the period up to 2011. Data for the city of Landskrona are presently digitized. The database contains a variety of information on individual as well as household/family level and each individual in the database is under observation from birth/in-migration and throughout the life span/until an out-migration occurs. The fact that, and the way in which it also combines economic and demographic data in one data base, have made it unique by Swedish comparisons. The main sources are: Church registers from 1646 (births, deaths, marriages), population registers from 1813/1829 (migration, household composition), poll-tax registers from 1697 (land ownership). Data on male heights from military records after 1830, data from mid-wife reports from 1880, and income from tax authorities from 1903 have been added amongst others. *Main goal* 1) To improve knowledge of individual behaviour and demographic outcomes during the agrarian society and during the transformation to the modern welfare society, 2) to enhance understanding of contemporary behaviour and health through a full life-course approach 3) to allow an analysis of the role of intergenerational transfers and inheritance on behaviour and 4) to analyse the influence of economic change and development of welfare institutions on individual behaviour.

5. Keywords: Please use the recommended keywords if they are applicable: *demography, life course, census, church register, civil certificates, population register, history, social science, genetics, migration, occupations.*	demography, life course, church register, population register, history, social science, migration, occupations
Please add your own keywords, if you have data not covered by the recommended terms.	Income and taxation registers, poll-tax registers, cause of death

6. **Citation:** Indicate how you want others to cite your database.	Bengtsson, T., M. Dribe, L. Quaranta and P. Svensson (2014). The Scanian Economic Demographic Database. Version 4.0 (Machine-readable database).

	Lund: Lund University, Centre for Economic Demography.

7. IDS compatible: Indicate with *Yes* or *No* whether the database is IDS compatible, if *Yes*, please specify.	Yes

8. Has the database already been completed or it is still under construction?	
8.a. If completed, please indicate the years of its construction?	5 parishes complete (1685-2010)
8.b. If under construction, please indicate, when it is planned to complete it?	1 city under construction
8.c. Please add a brief description of future plans for the database.	Complete registration of Landskrona

II. Contact information

1. Name of institute or organisation	Centre for Economic Demography, Lund university, Sweden
1.a. Website	http://www.ed.lu.se/databases/sedd
1.b. Location: city, country	Lund, Sweden
1.c. Postal address	Centre for Economic Demography, P.O. Box 7083, S-220 07 Lund, Sweden
1.d. Phone	

2. Name of primary responsible person	Martin Dribe
2.a. His/her email address	martin.dribe@ekh.lu.se
2.b. Postal address	Center for Economic Demography, PO Box 7083, 22007 Lund, Sweden
2.c. Phone	+46 46 2224677

3. Administrative information	
3.a. When this form was filled?	2015-03-10
3.b. Who did it?	Martin Dribe

4. Main economic funding (Name of organization(s) who made the grants /sustain it)	Swedish Research Council, Bank of Sweden Tercentenary Foundation

III. Sources: core characteristics

1. Type of the sources.
Indicate how many sources were used for the database and what kind (register, census, certificates …).
Please enter *Yes* or *No* and the time period for the main sources. In case of other sources, not listed below, please add their type and specify their main characteristics.
Detailed questions about the characteristics of all core sources are in section C.

	Type of source	***Yes/No***	***Start year***	***End year***	***Explanations:***
1.	Baptisms	Y	1646	1991	Or births: baptism for seventeenth century, birth date from late seventeenth century and onwards.
2.	Marriages from church registers	Y	1646	1991	
3.	Burials	Y	1646	1991	Or deaths: burial date up to mid-eighteenth century, death date from mid-eighteenth century and onwards

					(differences between parishes).
4.	Population registers, maintained by church	Y	1829	1991	
5.	Civil birth certificates	Y	1992		From 1992 and onwards.
6.	Civil marriage certificates	Y	1992		From 1992 and onwards.
7.	Civil death certificates	Y	1992		From 1992 and onwards.
8.	Population Census				
9.	Nominative lists				
10	Military draft records				
11.	Other: Civil population registers	Y	1968		From 1968 onwards
12.	Other: Poll tax registers	Y	1697	1945	
13.	Other: Land registers	Y	1658	1881	
14	Income and taxation registers	Y	1862		From 1862 onwards

IV. The database: core characteristics

1. Period covered by data: give first and last year of date, if possible	1646 – 2011
2. Territory covered by data	Scania
3. Geographical characteristic: local, regional, national, cross-national	

4. Units of observation. Please enter *Yes* or *No* for each unit, which forms the sample, the number of units and write explanations/comments. Add other units if they are not listed below, for them explanations are especially important.

	Units of observation:	**Yes /No**	**Number of units**	***Explanations:***
1.	Individuals		300,000	Estimate for the complete database, about 150,000 today
2.	Married couples			
3.	Families			
4.	Households			
5.	Farms			
6.	Institutions			
7.	Other			

5. Variables per unit included in the database	
On individuals: *Data of birth and dead, age, gender, marital status, religion, occupation, migration, relationship,* etc. Please add more variables, if they are not in the list	See metadata
On households: *Type of household, children present, age and number of children, etc*. Please add more variables, if they are not in the list	See metadata

6. Kinship relations:	
6.a. How is kinship recorded in the database?	Husband, wife and children are recorded in the examination and birth registers.
6.b. How deep (number of generations) is kinship information going?	3-4 generations can be analysed

7. **Completeness**	
7.a. Are all variables from the sources included in the database?	No
7.b. Are all individuals who lived in the households of	Yes

the sample recorded?	

8. Current data representation: Database Software (e.g. MySql, MsSql, Access, please specify	SQL server for data management, STATA for statistical analysis

9. Access conditions:	
9.a. How does a user get access to the database?	Data until 1910 available through home page, 1910- through application
9.b. What are the conditions and restrictions?	Compliance with ethical committee clearance 1910-

V. Publications and reports

1. Main publications about the database itself (max. 5)

2. Main or exemplary publications on research based on the database (max. 5)

- Bengtsson, T., et al. (2004). Life under Pressure. Mortality and Living Standards in Europe and Asia, 1700-1900. Cambridge: MIT Press.
- Bengtsson, T. & M. Dribe. (2006). Deliberate control in a natural fertility population: southern Sweden 1766-1865.Demography 43: 727-746
- Bengtsson, T. & M. Dribe. (2011). The late emergence of socioeconomic mortality differentials: A micro-level study of adult mortality in Southern Sweden 1815-1968. Explorations in Economic History 48: 389-400.
- Dribe, M. & C. Lundh (2010). Marriage choices and social reproduction. The interrelationship between partner selection and intergenerational socioeconomic mobility in 19th century Sweden.
- Demographic Research 22:347-382.
- Quaranta, L. (2013). Early life effects across the life course: The impact of individually defined exogenous measures of disease exposure on mortality by sex in 19th- and 20th-century Southern Sweden. Social Science & Medicine 119: 266-273

Section B

contains more specific and detailed questions about databases, such as the period(s) of observation, sampling design and procedures, data collection, linkage process and others.

VI. Observations

1. How do individuals enter observation?	By birth or in-migration, or present at start of database, e.g. 1646 or 1680.
2. How do individuals leave observation?	By death or out-migration, or at end of database period (presently 1968). After 1968 it is possible to follow individuals even after out-migration from the study area, why leaving observation is death, out-migration from Sweden, or end of database period (presently 2011).
3. How do households enter observation?	
4. How do households leave observation?	
5. Are some entry or exit dates unknown?	Yes, before 1829.
6. Are some entry or exit dates estimated?	Migration during the period 1829 to 1895 is set at specific dates in the autumn of each year. This occurs when an individual is registered as migrating in a certain year in a register and is not found in the same location (internal migration), or not at all (external migration) in the register of the following year. The same goes for family migration before 1829.
7. Can observations be linked to geographic locations?	Yes
8. Are the dates and locations of movements within the observation area recorded?	Partly, 1829 and onwards
9. Are all individuals who lived in selected households recorded? (Selection on basis of the sample or because sampled individuals are living in households)	Partly, 1829 and onwards
10. Are there related observations that are not included in the database?	a)For all ever married individuals residing in the parishes sometime between 1829 and 1968 information on socio-economic status at birth is registered whether they were born in the parishes or immigrants. This means that their parents' socio-economic status is registered although many lived outside the parishes. b) All marriages, which took place in the parishes, are recorded. This implies that the groom or the bride residing in another parish, but marrying in the parishes, is recorded even if the married couple settles down outside the parishes in the sample directly after marriage.

VII. Sampling design and procedures: how was sample(s) defined?

1. Source(s): Which source forms the basis for the sample	Population registers. No sample, all individuals in the registers are included
2. Sampling units: Households, individuals, regions…	See 1
3. Variables used for selection: Age, gender, marital status, other	See 1
4. Selection method: Random, stratified random, total count, clustered, other	All individuals residing in five rural parishes and one town in the province of Scania, southern Sweden. The sample is selected to represent all variations in economic and social structure on the countryside in Scania.

VIII. Data collection

1. Data collection period: When the data was collected and transcribed?	1984-present
2. Data collection method: Public digital register, transcription, other	Data entry from original sources
2.a. If transcription, how was the transcription done: ◦ By individuals ◦ From scanned sources ◦ From LDS's microfilms ◦ Automatic controls	By individuals
2.b. How was the checking of the transcription done? For example, by proof reading?	Semi-manual controls
2.c. When was it done?	Continuously
2.d. Purpose of the transcription: please indicate ◦ LDS ◦ Research ◦ Genealogy	Research
3. Control methods by researcher: e.g. Internal consistencies such as a death cannot happen before a birth of the same person	Many different controls of consistency
4. Data collection staff: Please indicate the number of people and their position (member of the project, free-lancer, other**)**	At present 6 data entry, 1 programmer/data manager

IX. Linkage process

1. Linkage: Which sources and units of observation have been linked: (e.g. birth/baptisms and death/burials…)?	Births/Baptisms -Y Marriages -Y Deaths/Burials -Y Population registers -Y Census - N Nominative lists -N Poll tax registers -Y Land registers -Y
2. Documentation of linking:	
2.a. Programme, manually, …	Manually
2.b. Name of software if used (and its parameters)	
3. What are the rules for linking? Flags definition (list them: age, name, extra knowledge …)	Name, date and place of birth, household contexts
4. How each reconstructed person is traceable to the original sources /transcribed data?	No systematic references in the data
5. How is linkage represented in the database? For example, do all occurrences of an individual include a universal identification number (ID)? Or are records linked in another way?	Each individual has a universal identification on number to which the different events are linked
6. Linkage percentage	NA
7. Quality of linkage (own evaluation)	NA
8. What reference/coding systems have been linked to the data? For example, occupational titles (like HISCO), locations (including geo-referenced systems). Please indicate the name of the system and how it was used. (Yes, No, Partly).	HISCO, HISCLASS, SOCPO, Cause of Death classification

Y/N/P	*Reference system*	*Explanations:*
Y	Occupational titles:	HISCO, and available in HISCLASS and SOCPO
Y	Locations (including geo-referenced systems):	Location by village and parish. Presently geo-coded locations on farm level are implemented for the 19th century.
N	Religion, civil status etc.:	
Y	Other:	Socio-economic status coding through the use of occupation and landholding. After 1903 personal income is also available.

Section C

contains detailed questions about sources used for the databases: their type, scope, content, state of preservation, etc.

Please answer the questions about all the sources used for the database, but do it in a separate form for every type of the source.

X. The main characteristics of the source (per every type of the source)

1. Official name of the source and its English translation	Husförhörslängd/församlingsbok (population register)
2. Purpose of the source:	
2.a. Why was this source created?	Administration
2.b. Who created it?	Clergy
3. Scope: What group of the population was documented in this source?	All registered individuals
4. Time period: When the information of the sources was recorded? Please indicate the start and the end date.	1813/29-1991 (for the parishes included)
5. Geographical area: What territory is covered by the source?	Nationwide
6. Content: What was recorded?	Demographic data, occupation, migration, household relations, vaccination (sometimes), catechetical examinations
7. Language of written material: original sources and documentation	Swedish
8. Preservation and storage:	
8.a. Completely preserved	Yes
8.b. Partially destroyed by personnel according to systematic criteria	
8.c. Partially destroyed or damaged for other reasons	
8.d. Reorganized by producer of the source	
8.e. Reorganized by record linkage procedures	
8.f. Where the original records are stored (name of the archive or institution)?	Regional archives, and municipal archives
9. Documentation:	
9.a. Completely documented and accessible by:	
9.b. Partially documented and accessible by:	Regional Archives, Lund, Nationella Arkivdatabasen (NAD, National Archives)
9.c. No documentation, but accessible by:	