

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

Documentation of the Databases of the EHPS-net

Improvement of the new Documentation of the databases on the EHPS website.
An ongoing report on the enhancement and updates of EHPS-Net.

v.1.3 (4/08/2017)

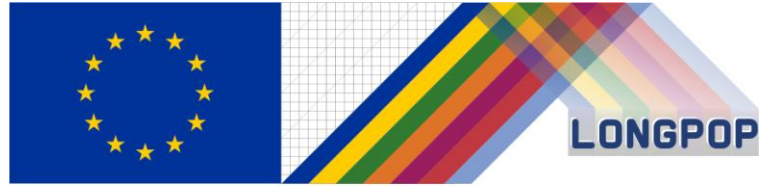
Francisco Anguita

Early-stage researcher of the LONGPOP project

This publication reflects only the author's view and the Agency is not responsible for any use that may be made of the information it contains

Table of Contents

2. Overview and Catalogue of the Databases in EHPS-net	3
3. China Multi-Generational Panel Dataset - Liaoning (CMGPD-LN)	4
3.1. Introduction	4
3.2. General Information.....	4
3.2.1 Contact Information.....	4
3.2.2. Type of Sources	4
3.2.3. Characteristics of the Database	4
3.2.4. Publications	5
3.3. Methodology.....	6
3.3.1. Periods of Observations.....	6
3.3.2. Sampling design and procedures: definition of sample(s).....	6
3.3.3. Data Collection	6
3.3.4. Linkage Process	6
3.4. Characteristics of the Sources.....	7
4. China Multi-Generational Panel Dataset - Shuangcheng (CMGPD-SC).....	7
4.1. Introduction	7
4.2 General Information.....	7
4.2.1 Contact Information.....	8
4.2.2. Type of Sources	8
4.2.3. Characteristics of the Database	8
4.2.4. Publications	8
4.3. Methodology.....	9
4.3.1. Periods of Observations.....	9
4.3.2. Sampling design and procedures: definition of sample(s).....	9
4.3.3. Data Collection	10
4.3.4. Linkage Process	10
4.4. Characteristics of the Sources.....	10
5. Appendix	11



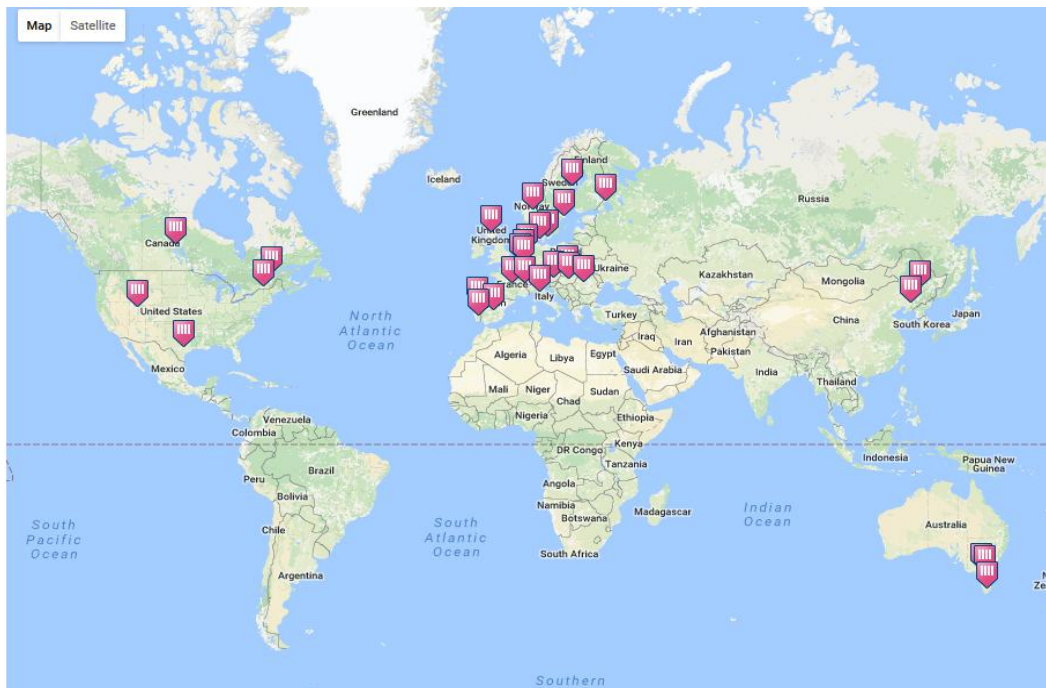
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

2. Overview and Catalogue of the Databases in EHPS-net.

The European Historical Population Samples Network (EHPS-Net) creates a portal that provides access to the European databases, as well as to important non-European ones which have joined the network. On the portal site an overview is given of all databases, their contents and the degree to which their variables have been structured to fit into the IDS.

The Network also brings together scholars to create a common format for databases containing non-aggregated information on persons, families and households. This format, the so called Intermediate Data Structure (IDS), aims to form an integrated and joint interface between many European databases in order to transcend the currently regional and national studies.

A list of all the databases included in the Network can be found in the appendix of this document (see page 11), but in this report we update the information of the two recently added new ones: The China Multi-Generational Panel Dataset-Liaoning (CMGPD-LN), and The China Multi-Generational Panel Dataset-Shuangcheng (CMGPD-SC).



3. China Multi-Generational Panel Dataset - Liaoning (CMGPD-LN)

3.1. Introduction

This database is drawn from the population registers compiled by the Imperial Household Agency (nei-wu-fu, 内务府) in Shengjing, currently the northeast Chinese province of Liaoning, between the years of 1749 and 1909. It provides 1.5 million triennial observations of more than 260,000 residents from 698 communities. The population mainly consists of immigrants from North China who settled in rural Liaoning during the early eighteenth century, and their descendants. The data provide socioeconomic, demographic, and other characteristics for individuals, households, and communities, and record demographic outcomes such as marriage, fertility, and mortality. The data also record specific disabilities for a subset of adult males. Additionally, the collection includes monthly and annual grain price data, custom records for the city of Yingkou, as well as information regarding natural disasters, such as floods, droughts, and earthquakes. This dataset is unique among publicly available population databases because of its time span, volume, detail, and completeness of recording, and because it provides longitudinal data not just on individuals, but on their +households, descent groups, and communities.

3.2. General Information

This database is not IDS compatible and it has been completed and publicly released. Its construction was finished between 1983 and 2010.

3.2.1 Contact Information

The name of the organization that developed it is HKUST whose web site is <http://www.ust.hk>, based in Clear Water Bay, Kowloon, Hong Kong, and being Cameron Campbell the contact person through email (camcam@ust.hk), postal address (Division of Social Science, HKUST, Clear Water Bay, Kowloon, Hong Kong) or phone (+852-2358-2776).

3.2.2. Type of Sources

The kind of sources employed for the database were population registers maintained by the church or the state between 1749 and 1909.

3.2.3. Characteristics of the Database

For this period of years stated in the previous paragraph, the territory covered was North and South-Central Liaoning. It is therefore, according to its geographical characteristics, a regional database, and its

units of observation are: 260.000 individuals, married couples (whose count depends on the definition), families, households, 29 state institutions where populations were organized, and 698 communities.

Some of the variables on individuals are: data of birth and dead, age, gender, marital status, religion, occupation, migration, relationship, presence or absence of various kin, official position (for males), banner affiliation, relationship to household head, number of children born, number of children alive. Also, variables for households inform of size and age composition.

In addition, kinship relations are recorded through patrilineal ascription, and it encompasses 6 generations depth.

About the database completeness, not all the variables from the sources are included. Names are only available in *pinyin*¹ and not the original Chinese characters. Also, all the individuals who lived in the household were recorded except for boys who died young, and daughters.

Concerning data representation, it is available as a delimited rectangular file, and in various formats for common statistical programs such as STATA, SPSS, SAS.

Finally, with respect to the access conditions, users can download the database from ICPSR, where some variables are freely downloadable and others -as *name*- requires some agreement signing.

3.2.4. Publications

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

- Lee, J.Z., Campbell, C.D., & Chen, S. (2010). *China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN) 1749-1909. User Guide*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research. [Download](#).
- Dong, H., Campbell, C., Kurosu, S., Yang, W., & Lee, J.Z. (2015). New sources for comparative social science: Historical population panel data from East Asia. *Demography*, 52 (3), 1061-1088. [Download](#).

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

- Lee, J., & Campbell, C. (1997). *Fate and Fortune in Rural China: Social Organization and Population Behavior in Liaoning, 1774-1873*. Cambridge University Press.
- Xi, S., Campbell, C., & Lee, J.Z. (2015). Ancestry Matters: Patrilineage Growth and Extinction. *American Sociological Review*, 80(3), 574-602. [Download](#).
- Cameron, C., & Lee, J.Z. (2011). Kinship and the Long-Term Persistence of Inequality in Liaoning, China, 1749-2005. *Chinese Sociological Review*, 44(1), 71-104. PMID: PMC23596557.

¹ is the official Romanization system for Standard Chinese in mainland China.

- Tsuya, N. O., Campbell, C., & Feng, W. (2010). Reproduction: Models and Sources. Chapter 3 in Noriko Tsuya, Wang Feng, George Alter, James Z. Lee et al. *Prudence and Pressure: Reproduction and Human Agency in Europe and Asia, 1700-1900*. MIT Press, 39-64.
- Campbell, C., & Lee, J. (2008). Kin Networks, Marriage, and Social Mobility in Late Imperial China. *Social Science History*, 32(2), 175-214.

3.3. Methodology

3.3.1. Periods of Observations

Individuals enter observation in a triennial register and leave it annotating departure. In turn, households are recorded in cross section every three years.

Some entry and exit dates are unknown and some others are estimated. Also, observations can be linked to geographic locations. Dates and locations of movements within the observation area are recorded. In addition, all individuals who lived in a selected household are recorded, except for daughters and boys who died early. There are related observations that were not included in the database (additional registers from the region that were not coded).

3.3.2. Sampling design and procedures: definition of sample(s)

The sources that form the basis of the sample are household registers compiled every three years for populations affiliated with the Eight Banners in Liaoning, including tenants on Eight Banner land. In the selection method 29 sets of registers were entered in their entirety. These were selected from available sets based on their completeness. Other sets had more gaps between extent registers.

3.3.3. Data Collection

Data was collected between 1983 and 2010 as transcriptions from microfilm and scanned images. This was performed by individuals from scanned sources, including LDS microforms and scans and after consistency checking of the variables.

3.3.4. Linkage Process

Manual linking was undertaken for individuals across triennial registers, and programmed linkage with STATA was applied to these individuals to their kin. The sources that were linked are longitudinal triennial registers, and the units of observation are family members and ancestry.

The rules applied for linking were: name, location in register, age and other information not specified. Linkage is represented through a unique identifier assigned to all individuals and its percentage achieved is nearly complete².

Finally, latitude and longitude are the reference system for locations linked to the data.

3.4. Characteristics of the Sources

The Eight Banner system created this source in order to track tenants on land associated with the Eight Banner system and other populations affiliated with the Eight Banners. The groups of population documented in this database were the administrative populations in Liaoning affiliated with the Eight Banner system during the period of 1749 and 1909, covering the northeast, central, and south-central Liaoning geographical areas. Its content –the variables recorded- were: Name, age, relationship to household head, official position, and other details not specified.

The original source was written in Chinese and it was completely preserved.

The sources are stored in the Liaoning Provincial Archives, LDS has microforms/scans.

4. China Multi-Generational Panel Dataset - Shuangcheng (CMGPD-SC)

4.1. Introduction

provides longitudinal individual, household, and community information on the demographic and socioeconomic characteristics of a resettled population living in Shuangcheng. This is a county in present-day Heilongjiang Province of Northeastern China, and it includes registers for the period from 1866 to 1913. The dataset includes around 1.3 million annual observations of over 100,000 unique individuals descended from families who were relocated to Shuangcheng in the early 19th century. These families were divided into 3 categories based on their place of origin: metropolitan bannermen, rural bannermen, and floating bannermen. The CMGPD-SC, like its Liaoning counterpart, the CMGPD-LN (ICPSR 27063), is a valuable data source for studying longitudinal as well as multi-generational social and demographic processes. The population categories had salient differences in social origins and land entitlements, and landholding data are available at a number of time periods, thus the CMGPD-SC is especially suitable to the study of stratification processes.

4.2 General Information.

This database is not IDS compatible and it has been completed and publicly released. Its construction was finished between 2004 and 2015.

² This is why the quality of linkage is self-proclaimed as “awesome”.

4.2.1 Contact Information

The name of the organization that developed it is HKUST (<http://www.ust.hk>), based in Clear Water Bay, Kowloon, Hong Kong. Shuang Chen is the contact person via email (camcam@ust.hk), postal address (Department of History 280 Schaeffer Hall University of Iowa Iowa City, Iowa 52242-1409) or phone (+852-2358-2776).

4.2.2. Type of Sources

The type of sources employed for the database were population registers maintained by the church or the state between 1866 and 1913, but also, landholding records.

4.2.3. Characteristics of the Database

For the period of years already mentioned (section 2.2.2)., the territory covered was Shuangcheng County in Heilongjiang, China. It is therefore, according to its geographical characteristics, a regional database, and its units of observation are 100,000 individuals, married couples (which count depends on the definition), households, 8 institutions where populations was organized by Banner, and 120 communities.

Some of the variables on individuals are: data of birth and dead, year of marriage, marital status, presence or absence of various kin, official position (for males), banner affiliation, age (in Chinese sui), relationship to household head, number of children born, number of children alive, ethnic registration, nature of institutional affiliation, landholding (via linkage to land registers). Also, variables for households inform of size, age composition and landholding.

In addition, kinship relations are recorded through multiple variables available under patrilineal ascription, and it is as deep as 4 generations.

About the database completeness, not all the variables from the sources are included. Names are only available in pinyin and not the original Chinese characters. Also, all the individuals who lived in the household were recorded except for boys who died young, and daughters.

Concerning data representation, it is available as a delimited rectangular file, and in various formats for common statistical programs such as STATA, SPSS, SAS.

Finally, with respect to the access conditions, users can download the database from ICPSR, where some variables are freely downloadable and others -as *name*- requires some agreement signing.

4.2.4. Publications

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

- Lee, J.Z., Campbell, C.D., & Chen, S. (2010). *China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN) 1749-1909. User Guide*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research. [Download](#).
- Dong, H., Campbell, C., Kurosu, S., Yang, W., & Lee, J.Z. (2015). New sources for comparative social science: Historical population panel data from East Asia. *Demography*, 52 (3), 1061-1088. [Download](#).

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

- Lee, J., & Campbell, C. (1997). *Fate and Fortune in Rural China: Social Organization and Population Behavior in Liaoning, 1774-1873*. Cambridge University Press.
- Xi, S., Campbell, C., & Lee, J.Z (2015). Ancestry Matters: Patrilineage Growth and Extinction. *American Sociological Review*, 80(3), 574-602. [Download](#).
- Cameron, C., & Lee, J.Z. (2011). Kinship and the Long-Term Persistence of Inequality in Liaoning, China, 1749-2005. *Chinese Sociological Review*, 44(1), 71-104. PMID: PMC23596557.
- Tsuya, N. O., Campbell, C., & Feng, W. (2010). Reproduction: Models and Sources. Chapter 3 in Noriko Tsuya, Wang Feng, George Alter, James Z. Lee et al. *Prudence and Pressure: Reproduction and Human Agency in Europe and Asia, 1700-1900*. MIT Press, 39-64.
- Campbell, C., & Lee, J. (2008). Kin Networks, Marriage, and Social Mobility in Late Imperial China. *Social Science History*, 32(2), 175-214.

4.3. Methodology

4.3.1. Periods of Observations

Individuals enter observation when they appear in an annual register, and they leave when their departure is annotated in an annual register.

Some entry and exit dates are unknown and some others are estimated. Also, observations can be linked to geographic locations. Dates and locations of movements within the observation area are recorded. In addition, all individuals who lived in a selected household are recorded, except for daughters and boys who died early. There are related observations that were not included in the database (some registers have not survived).

4.3.2. Sampling design and procedures: definition of sample(s)

The sources that form the basis of the sample are household registers compiled every year for a settler population in Shuangcheng. The sampling units were individuals and the variables employed for the selection were all available.

Finally, the selection methods were all that were available and achieving the entering of 12 sets of registers in their entirety.

4.3.3. Data Collection

Data was collected between 2005 and 2015 as transcriptions from microfilm and scanned images. This was performed by individuals from scanned sources, including LDS microforms and scans and after consistency checking of the variables.

The purpose of the transcription was research, though the name index was turned over to the LDS. Also, numerous consistency checks were carried out along these lines.

4.3.4. Linkage Process

Manual linking was undertaken for individuals across triennial registers, and programmed linkage with STATA was applied to these individuals to their kin. The sources that were linked are longitudinal annual registers, and the units of observation are family members and ancestry.

The rules applied for linking were: name, location in register, age and other information not specified. Linkage is represented through a unique identifier assigned to all individuals and its percentage achieved is nearly complete³.

Finally, concerning reference systems, no occupation titles were linked to the data although latitude and longitude are as for locations.

4.4. Characteristics of the Sources

This source was created by the Eight Banner system in order to track settlers and their descendants in Shuangcheng. The groups of population documented in this database were the administrative populations located in the above mentioned geographical region⁴ and affiliated with the Eight Banner system. The information of the sources was recorded during the period between 1866 and 1913, although the registration started probably earlier but those registers did not survive.

Its content –the variables recorded- were: name, age, relationship to household head, official position, and other details not specified. Landholding was registered in other registers, which has also been entered.

The original source was written in Chinese and it was completely preserved.

The sources are stored in the Liaoning Provincial Archives, LDS has microforms/scans.

³ This is why the quality of linkage is self-proclaimed as “amazing”.

⁴ Shuangcheng

5. Appendix

Namesort descending	Code	From	To	Territory
Antwerp COR*-database	COR*	1846	1920	Antwerp district, Belgium
Aranjuez Database: Individual and family trajectories	ADB	1871	1970	Aranjuez, Spain
BALSAC: Quebec population database, 1621-1971	BALSAC	1621	1971	Québec, Canada
Base TRA Patrimoine	TRA	1800	1960	France
China Multigenerational Panel Database-Liaoning	CMGPD-LN	1749	1909	North and south-central Liaoning, China
China Multigenerational Panel Database-Shuangcheng	CMGPD-SC	1866	1913	Shuangcheng County, Heilongjiang, China
Female Demographic Biographies: Wald parish, 1880-1938	FDB-Wald	1880	1938	Wald parish (Catholic and Lutheran part)
Founders & Survivors: Tasmanian life courses in historical context	FAS	1803	1930	Australian colonies and states, United Kingdom and Ireland
Geneva Demographic Database	GDD	1800	1880	Geneva, Switzerland
Historical Database of the Liège Region	HDLR	1806	1900	Liège, Belgium
Historical population database of Transylvania, 1850-1914	HPDT	1850	1914	Transylvania, Romania
Historical Sample of the Netherlands	HSN	1812	1922	The Netherlands
Historical Sample Portuguese Social Mobility, 1850-1960	HSPM	1860	1957	Portugal
Hungarian Historical Demographic Database	HHDD	1800	1945	Roman Catholic parishes of Szentegyházásfalva and Kápolnásfalva in Eastern Transylvania (Romania).
Integral History Project Groningen	IHPG	1770	1914	Groningen, Netherlands
Italian Historical Population Database	IHPD	1796	1883	Casalguidi (Tuscany) and Madregolo (Emilia), Italy
Karelian Database	KATIHA	1681	1950	The Old Eastern Finland, Karelia
Koori Health Research Database	KHRD	1855	1930	Australia
Melbourne Lying-In Hospital Cohort	MLHC	1803	1985	Australia
Mosaic project	MOSAIC	1430	1918	Europe, best coverage until now for Central and East Central Europe
National Sample of the 1901 Census of Canada	CFP	1901	1901	Canada
Norwegian Historical Population Register, 1800-1964	HPR	1800	1964	Norway
Odense database: Persons and buildings in Odense, 1741-1921	ODENSE	1761	1900	Odense, Denmark
POPLINK	DDB	1900	1950	Skellefteå region, Umeå region, Sweden
POPUM	DDB, POPUM	1620	1900	Sweden
Portuguese Genealogical Repository	PORGENER	1550	1910	Portugal
Registre de la population du Québec ancien	RPQA	1621	1850	Quebec, Canada
Scanian Economic Demographic Database	SEDD	1646	2011	Scania, Sweden
Texas Counties Database	TCD	1850	1910	Texas, US
The Demography of Victorian Scotland: Linked data for 4 Scottish communities, 1861-1901	DVS	1860	1901	Scotland
The Roteman Database	TRD	1878	1926	Stockholm, Sweden
Utah Population Database	UPDB	1790	present	Utah, US