SURVEY OF HISTORICAL DATABASES

WITH LONGITUDINAL MICRO-DATA

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	China Multigenerational Panel Database-Liaoning
1.a. Subtitle, which brings meaning to the title (scope, place, time period):	Liaoning, China, 1749-1909
2. Abbreviation	CMGPD-LN

3. Links to website(s):	
3.a. Homepage	http://www.icpsr.umich.edu/icpsrweb/DSDR/studies/27063
3.b. Get to data	Same as above

| 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 China Multi-Generational Panel Dataset - Liaoning (CMGPD-LN) is drawn from the population registers compiled by the Imperial Household Agency (neiwufu) in Shengjing, currently the northeast Chinese province of Liaoning, between 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.

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.*	**History, demography, population register**
Please add your own keywords, if you have data not covered by the recommended terms.	**China, multi-generational**

6. **Citation:** Indicate how you want others to cite your database.	Lee, James Z., and Cameron D. Campbell. China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN), 1749-1909. ICPSR27063-v10. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2016-09-06. http://doi.org/10.3886/ICPSR27063.v10

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

8. Has the database already been completed or it is still under construction?	Completed and publicly released
8.a. If completed, please indicate the years of its construction?	1983-2010
8.b. If under construction, please indicate, when it is planned to complete it?	
8.c. Please add a brief description of future plans for the database.	

II.   **Contact information**

1. Name of institute or organisation	**HKUST**
1.a. Website	**http://www.ust.hk**
1.b. Location: city, country	**Clear Water Bay, Kowloon, Hong Kong**
1.c. Postal address	
1.d. Phone	

2. Name of primary responsible person	**Cameron Campbell**
2.a. His/her email address	**camcam@ust.hk**
2.b. Postal address	**Division of Social Science, HKUST, Clear Water Bay, Kowloon, Hong Kong**
2.c. Phone	**+852-2358-2776**

3. Administrative information	
3.a. When this form was filled?	**22 September 2016**

3.b. Who did it?	Cameron Campbell

4. Main economic funding (Name of organization(s) who made the grants /sustain it)	United States Department of Health and Human Services. National Institutes of Health. Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01 HD057175-01A1)

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				
2.	Marriages from church registers				
3.	Burials				
4.	Population registers, maintained by church or state	Yes	1749	1909	
5.	Civil birth certificates				
6.	Civil marriage certificates				
7.	Civil death certificates				
8.	Population Census				
9.	Nominative lists				
10	Military draft records				
11.	Other:				

IV. The database: core characteristics

1. Period covered by data: give first and last year of date, if possible	**1749-1909**
2. Territory covered by data	**North and south-central Liaoning**
3. Geographical characteristic: local, regional, national, cross-national	**Regional**

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	Yes	260,000	
2.	Married couples	Yes		Counts depend on definitions
3.	Families	Yes		
4.	Households	Yes		
5.	Farms			
6.	Institutions	Yes	29	Populations were organized by state institution
7.	Other	Yes	698	Communities

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	Year of birth (calculated) Year of death (3 year interval) Year of marriage (3 year interval) Marital status (inferred)

3

	Presence or absence of various kin Official position (for males) Banner affiliation Age (in Chinese sui) Relationship to household head Numbers of children born Numbers of children alive
On households: *Type of household, children present, age and number of children, etc.* Please add more variables, if they are not in the list	Size Age composition

6. Kinship relations:	
6.a. How is kinship recorded in the database?	Patrilineal kinship (multiple variables available)
6.b. How deep (number of generations) is kinship information going?	6 generations

7. **Completeness**	
7.a. Are all variables from the sources included in the database?	No. Names are only available in pinyin, not the original Chinese characters.
7.b. Are all individuals who lived in the households of the sample recorded?	All except for boys who died young, and daughters.

8. Current data representation: Database Software (e.g. MySql, MsSql, Access, please specify	Available as a delimited rectangular file, and in various formats for common statistical programs such as STATA, SPSS, SAS

9. Access conditions:	
9.a. How does a user get access to the database?	Download from ICPSR
9.b. What are the conditions and restrictions?	Certain variables, such as name require signing an agreement with ICPSR. Other variables may all be downloaded freely.

V.   Publications and reports

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

Lee, James Z., Cameron D. Campbell, and Shuang Chen. 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 at: http://www.icpsr.umich.edu/cgi-bin/file?comp=none&study=27063&ds=0&file_id=1222828&path=DSDR

Dong, H., Campbell, C., Kurosu, S., Yang, W., Lee, J.Z. . New sources for comparative social science: Historical population panel data from East Asia. Demography. 52, (3), 1061-1088.

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

Lee, James and Cameron Campbell. 1997. *Fate and Fortune in Rural China: Social Organization and Population Behavior in Liaoning, 1774-1873.* Cambridge University Press.

SONG Xi, Cameron Campbell, and James Z. Lee. 2015. "Ancestry Matters: Patrilineage Growth and Extinction." *American Sociological Review*. 80(3):574–602 http://asr.sagepub.com/content/80/3/574

Campbell, Cameron and James Z. Lee. 2011. "Kinship and the Long-Term Persistence of Inequality in Liaoning, China, 1749-2005." *Chinese Sociological Review*. 44(1):71-104. PMCID: PMC23596557.

Tsuya, Noriko O., Cameron Campbell, and Wang Feng. 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, Cameron and James Lee. 2008. "Kin Networks, Marriage, and Social Mobility in Late Imperial China." *Social Science History*. 32(2):175-214.

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?	**Appear in a triennial register.**
2. How do individuals leave observation?	**Departure annotated in a triennial register.**
3. How do households enter observation?	**Recorded in cross-section every three years.**
4. How do households leave observation?	
5. Are some entry or exit dates unknown?	**Yes**
6. Are some entry or exit dates estimated?	**Yes**
7. Can observations be linked to geographic locations?	**Yes**
8. Are the dates and locations of movements within the observation area recorded?	**Yes**
9. Are all individuals who lived in selected households recorded? (Selection on basis of the sample or because sampled individuals are living in households)	Yes, except for daughters, and boys who died early.
10. Are there related observations that are not included in the database?	**There are additional registers from the region that we have not coded.**

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

1. Source(s): Which source forms the basis for the sample	Household registers compiled every three years for populations affiliated with the Eight Banners in Liaoning, including tenants on Eight Banner land.
2. Sampling units: Households, individuals, regions…	Individuals
3. Variables used for selection: Age, gender, marital status, other	All available.
4. Selection method: Random, stratified random, total count, clustered, other	All available. We entered 29 sets of registers in their entirety. These were selected from available sets based on their completeness. Other sets had more gaps between extent registers.

VIII.   Data collection

1. Data collection period: When the data was collected and transcribed?	1983-2010
2. Data collection method: Public digital register, transcription, other	Transcription from microfilm and scanned images.
2.a. If transcription, how was the transcription done: ◦ By individuals	By individuals from scanned sources, including LDS microforms and scans.

◦ From scanned sources ◦ From LDS's microfilms ◦ Automatic controls	
2.b. How was the checking of the transcription done? For example, by proof reading?	Consistency checking of the variables.
2.c. When was it done?	Continually
2.d. Purpose of the transcription: please indicate ◦ LDS ◦ Research ◦ Genealogy	Research, though we turned our name index over to the LDS.
3. Control methods by researcher: e.g. Internal consistencies such as a death cannot happen before a birth of the same person	Numerous consistency checks along these lines.
4. Data collection staff: Please indicate the number of people and their position (member of the project, free-lancer, other**)**	Coders hired on an ad hoc basis.

IX. Linkage process

1. Linkage: Which sources and units of observation have been linked: (e.g. birth/baptisms and death/burials…)?	Triennial registers, longitudinally. Family members and ancestry.
2. Documentation of linking:	
2.a. Programme, manually, …	Manual linkage of individuals across triennial registers, programme linkage of individuals to their kin.
2.b. Name of software if used (and its parameters)	STATA
3. What are the rules for linking? Flags definition (list them: age, name, extra knowledge …)	Name, location in register, age, other information.
4. How each reconstructed person is traceable to the original sources /transcribed 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?	All individuals are assigned a unique identifier.
6. Linkage percentage	Nearly complete.
7. Quality of linkage (own evaluation)	Awesome.
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).	

Y/N/P	*Reference system*	*Explanations:*
	Occupational titles:	None
	Locations (including geo-referenced systems):	Latitude/longitude
	Religion, civil status etc.:	
	Other:	

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 <u>separate form</u> 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	
2. Purpose of the source:	
2.a. Why was this source created?	**Track tenants on land associated with the Eight Banner system, and other populations affiliated with the Eight Banners**
2.b. Who created it?	**The Eight Banner system**
3. Scope: What group of the population was documented in this source?	**Administrative populations in Liaoning affiliated with the Eight Banner system**
4. Time period: When the information of the sources was recorded? Please indicate the start and the end date.	**1749-1909**
5. Geographical area: What territory is covered by the source?	**Northeast, central, and south-central Liaoning**
6. Content: What was recorded?	**Name, age, relationship to household head, official position, and other details.**
7. Language of written material: original sources and documentation	**Chinese**
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)?	**Liaoning Provincial Archives, LDS has microforms/scans**
9. Documentation:	
9.a. Completely documented and accessible by:	
9.b. Partially documented and accessible by:	
9.c. No documentation, but accessible by:	