

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:

I.	A	General (identifying) information about databases
II.	A	Contact information
III.	A	Sources: core characteristics
IV.	A	Database: core characteristics
V.	A	Publications and reports
VI.	B	Observations
VII.	B	Sampling design and procedures
VIII.	B	Data collection
IX.	B	Linkage process
X.	C	Sources: the main characteristics of every source

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-Shuangcheng
1.a. Subtitle , which brings meaning to the title (scope, place, time period):	Shuangcheng County, Heilongjiang Province, China, 1866-1913
2. Abbreviation	CMGPD-SC

3. Links to website(s):	
3.a. Homepage	http://www.icpsr.umich.edu/icpsrweb/DSDR/studies/35292
3.b. Get to data	Same as above

<p>4. Abstract: describes content of the database. Max. length: 300 words Please indicate:</p> <ul style="list-style-type: none"> ◦ Scope and main goal ◦ Time and territory covered by data ◦ Sample strategy ◦ Main sources 	<p>The China Multi-Generational Panel Dataset - Shuangcheng (CMGPD-SC) provides longitudinal individual, household, and community information on the demographic and socioeconomic characteristics of a resettled population living in Shuangcheng, a county in present-day Heilongjiang Province of Northeastern China, for the period from 1866 to 1913. The dataset includes some 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</p>
--	---

	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.
--	--

5. Keywords: Please use the recommended keywords if they are applicable: <i>demography, life course, census, church register, civil certificates, population register, history, social science, genetics, migration, occupations.</i>	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 <i>Yes</i> or <i>No</i> whether the database is IDS compatible, if <i>Yes</i> , 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?	2004-2015
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	Shuang Chen
2.a. His/her email address	shuang-chen@uiowa.edu
2.b. Postal address	Department of History 280 Schaeffer Hall University of Iowa Iowa City, Iowa 52242-1409
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)	Preparation of the CMGPD-SC and accompanying documentation for public release via DSDR at ICPSR was supported by the National Institutes of Health, Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) Grant no. R01 HD070985 “Multi-generational Demographic and Landholding Data: CMGPD-SC Public Release.”
---	--

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 <i>Yes</i> or <i>No</i> 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.					
	<i>Type of source</i>	<i>Yes/No</i>	<i>Start year</i>	<i>End year</i>	<i>Explanations:</i>
1.	Baptisms				
2.	Marriages from church registers				
3.	Burials				
4.	Population registers, maintained by church or state	Yes	1866	1913	
5.	Civil birth certificates				
6.	Civil marriage certificates				
7.	Civil death certificates				
8.	Population Census				
9.	Nominative lists				
10.	Military draft records				
11.	Other:	Yes			Landholding records

IV. The database: core characteristics

1. Period covered by data: give first and last year of date, if possible				1866-1913
2. Territory covered by data				Shuangcheng County, Heilongjiang, China
3. Geographical characteristic: local, regional, national, cross-national				Regional
4. Units of observation. Please enter <i>Yes</i> or <i>No</i> 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	100,000	
2.	Married couples	Yes		Counts depend on definitions
3.	Families	Yes		
4.	Households	Yes		
5.	Farms			
6.	Institutions	Yes	8	Populations were organized by Banner
7.	Other	Yes	120	Communities

5. Variables per unit included in the database	
<p><u>On individuals:</u> <i>Data of birth and dead, age, gender, marital status, religion, occupation, migration, relationship, etc.</i> Please add more variables, if they are not in the list</p>	<p>Year of birth (calculated) Year of death (3 year interval) Year of marriage (3 year interval) Marital status (inferred) 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 Ethnic registration Nature of institutional affiliation Landholding (via linkage to land registers)</p>
<p><u>On households:</u> <i>Type of household, children present, age and number of children, etc.</i> Please add more variables, if they are not in the list</p>	<p>Size Age composition Landholding</p>
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?	4 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)

Wang, Hongbo, Shuang Chen, Hao Dong, Matt Noellert, Cameron Campbell, and James Z Lee. 2013. China Multi-Generational Panel Dataset, Shuangcheng (CMGPD-SC) 1866-1913. User Guide. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

http://www.icpsr.umich.edu/cgi-bin/file?comp=none&study=35292&ds=0&file_id=1207512&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)

Chen Shuang, Cameron Campbell, and James Lee. 2014. "Categorical Inequality and Gender Difference: Marriage and Remarriage in Northeast China, 1749-1913." Chapter 11 in Lundh, Christer, Satomi Kurosu, et al. *Similarity in Difference: Marriage in Europe and Asia, 1700-1900*. MIT Press, 393-438.

CHEN Shuang, James Lee, and Cameron Campbell. 2010. "Wealth stratification and reproduction in Northeast China, 1866-1907." *History of the Family*. 15:386-412. PMID: PMC2992971.

任玉雪 (Ren Yuxue, 李中清 (James Lee), 康文林(Cameron Campbell). 2011 (民国 100 年). 地方政府實踐與國家制度之間的衝突及重塑-以晚清吉林將軍雙城堡民界的出現為列 (Conflict and Reform in Local Government Administrative Practices and Central Government Institutions: The Emergence of the Minjie in Shuangchengpu during the Late Qing Dynasty). *中央研究院历史语言研究所集刊 (Journal of the Academia Sinica Institute of History and Philology)* . 82(3): 493-532.

Shuang Chen, Cameron Campbell, and James Lee. 2005 (publ. 2006). "Vulnerability and Resettlement: Mortality Differences in Northeast China by Place of Origin, 1870-1912 - Comparing Urban and Rural Migrants." *Annales de Démographie Historique*. 2005(2): 47-79.

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 an annual register.
2. How do individuals leave observation?	Departure annotated in an annual register.
3. How do households enter observation?	Recorded in cross-section every year.
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?	Some registers have not survived.

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 year for a settler population in Shuangcheng.
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 12 sets of registers in their entirety.
---	---

VIII. Data collection

1. Data collection period: When the data was collected and transcribed?	2005-2015
2. Data collection method: Public digital register, transcription, other	Transcription from microfilm and scanned images.
2.a. If transcription, how was the transcription done: <ul style="list-style-type: none"> ◦ By individuals ◦ From scanned sources ◦ From LDS's microfilms ◦ Automatic controls 	By individuals from scanned sources, including LDS microforms and scans.
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 <ul style="list-style-type: none"> ◦ 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...)?	Annual registers, longitudinally. Family members and ancestry.
2. Documentation of linking:	
2.a. Programme, manually, ...	Manual linkage of individuals across 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)	Amazing.
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	<i>Reference system</i>	<i>Explanations:</i>
	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 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	
2. Purpose of the source:	
2.a. Why was this source created?	Track settlers and their descendants in Shuangcheng.
2.b. Who created it?	The Eight Banner system
3. Scope: What group of the population was documented in this source?	Administrative populations in Shuangcheng affiliated with the Eight Banner settlement.
4. Time period: When the information of the sources was recorded? Please indicate the start and the end date.	1866-1913. Registration probably started earlier, but the registers do not survive.
5. Geographical area: What territory is covered by the source?	Shuangcheng.
6. Content: What was recorded?	Name, age, relationship to household head, official position, and other details. Landholding was recorded in other registers, which have also been entered.
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:	