

Historical Chinese Microdata. 40 Years of Dataset Construction by the Lee-Campbell Research Group

By Cameron D. Campbell and James Z. Lee

To cite this article: Campbell C. D. & Lee, J. Z. (2020). Historical Chinese Microdata. 40 Years of Dataset Construction by the Lee-Campbell Research Group. *Historical Life Course Studies*, Online first. <http://hdl.handle.net/10622/23526343-2020-0004?locatt=view:master>

HISTORICAL LIFE COURSE STUDIES

Content, Design and Structure of Major Databases with
Historical Longitudinal Population Data

VOLUME 10, SPECIAL ISSUE 1,
2020

GUEST EDITORS

Sören Edvinsson
Kees Mandemakers
Ken Smith



MISSION STATEMENT

HISTORICAL LIFE COURSE STUDIES

Historical Life Course Studies is the electronic journal of the *European Historical Population Samples Network* (EHPS-Net). The journal is the primary publishing outlet for research involved in the conversion of existing European and non-European large historical demographic databases into a common format, the Intermediate Data Structure, and for studies based on these databases. The journal publishes both methodological and substantive research articles.

Methodological Articles

This section includes methodological articles that describe all forms of data handling involving large historical databases, including extensive descriptions of new or existing databases, syntax, algorithms and extraction programs. Authors are encouraged to share their syntaxes, applications and other forms of software presented in their article, if pertinent, on the EHPS-Net website.

Research articles

This section includes substantive articles reporting the results of comparative longitudinal studies that are demographic and historical in nature, and that are based on micro-data from large historical databases.

Historical Life Course Studies is a no-fee double-blind, peer-reviewed open-access journal supported by the European Science Foundation (ESF, <http://www.esf.org>), the Scientific Research Network of Historical Demography (FWO Flanders, <http://www.historicaldemography.be>) and the International Institute of Social History Amsterdam (IISH, <http://socialhistory.org/>). Manuscripts are reviewed by the editors, members of the editorial and scientific boards, and by external reviewers. All journal content is freely available on the internet at <http://www.ehps-net.eu/journal>.

Co-Editors-In-Chief:

Paul Puschmann (Radboud University) & Luciana Quaranta (Lund University)
hislives@kuleuven.be

The European Science Foundation (ESF) provides a platform for its Member Organisations to advance science and explore new directions for research at the European level. Established in 1974 as an independent non-governmental organisation, the ESF currently serves 78 Member Organisations across 30 countries. EHPS-Net is an ESF Research Networking Programme.



The European Historical Population Samples Network (EHPS-net) brings together scholars to create a common format for databases containing non-aggregated information on persons, families and households. The aim is to form an integrated and joint interface between many European and non-European databases to stimulate comparative research on the micro-level.
Visit: <http://www.ehps-net.eu>.



HISTORICAL LIFE COURSE STUDIES
ONLINE FIRST (2020), published 10-09-2020

Historical Chinese Microdata

40 Years of Dataset Construction by the Lee-Campbell Research Group

Cameron D. Campbell

Hong Kong University of Science and Technology & Central China Normal University

James Z. Lee

Hong Kong University of Science and Technology

ABSTRACT

The Lee-Campbell Group has spent forty years constructing and analysing individual-level datasets based largely on Chinese archival materials to produce a scholarship of discovery. Initially, we constructed datasets for the study of Chinese demographic behaviour, households, kin networks, and socioeconomic attainment. More recently, we have turned to the construction and analysis of datasets on civil and military officials and other educational and professional elites, especially their social origins and their careers. As of July 2020, the datasets include nominative information on the behaviour and life outcomes of approximately two million individuals. This article is a retrospective on the construction of these datasets and a summary of their findings. This is the first time we have presented all our projects together and discussed them and the results of our analysis as a single integrated whole. We begin by summarizing the contents, organization, and notable features of each dataset and provide an integrated history of our data construction, starting in 1979 up to the present. We then summarize the most important results from our research on demographic behaviour, family, and household organization, and more recently inequality and stratification. We conclude with a reflection on the importance of data discovery, flexibility, interaction and collaboration to the success of our efforts.

Keywords: China, Historical demography, Historical big data, Quantitative history, Inequality

e-ISSN: 2352-6343

PID article: <http://hdl.handle.net/10622/23526343-2020-0004?locatt=view:master>

The article can be downloaded from [here](#).

© 2020, Campbell, Lee

This open-access work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/), which permits use, reproduction & distribution in any medium for non-commercial purposes, provided the original author(s) and source are given credit. See <http://creativecommons.org/licenses/>.

1 INTRODUCTION

For over forty years, beginning in 1979, the Lee-Campbell Group has devoted considerable effort to locate, construct, and analyse individual-level datasets based largely on Chinese archival materials to produce a scholarship of discovery. Initially, we studied Chinese demographic behaviour, households, kin networks, and socioeconomic attainment. We constructed datasets that followed individuals from birth to death and families and households across multiple generations. More recently, we have turned to the study of the social origins and careers of civil and military officials and other educational and professional elites. We are now constructing datasets that describe the civil service careers of Qing dynasty officials from the 18th century to the beginning of the 20th century, the social origins and educational trajectories of university students during the 20th century, the qualifications and careers of government officials and educated professionals largely during the Republican era (1911–1949), and the experience of hundreds of thousands of Chinese peasants and their families during the process of rural reconstruction from land reform in the mid-1940s to Peoples Communes in the mid-1960s. Based on these data, we have published seven academic books and some 70 academic articles, mostly in English, eleven of which have won thirteen best academic prizes or equivalent recognition, including four books and two articles we wrote in English and two books and three articles we wrote in Chinese.¹

This article is a retrospective on these projects and a summary of their findings. In part one, we overview the datasets themselves, summarizing their contents, organization, and notable features. In part two, we provide an integrated history, starting in 1979 with James Lee's effort to locate systematic historical demographic microdata in China, and continuing up to the present. In part three, we summarize the contributions from the analysis of these datasets beginning with the key demographic outcomes that were the focus of our early work, and then move to inequality and stratification. We conclude with reflections based on our experience.

This is the first time we have presented all our projects together and discussed them and the results of our analysis as a single integrated whole. By doing so, we clarify the full scale and range of our efforts for readers who may only be familiar with some of the projects. The projects share an emphasis on the discovery of social phenomena by an inductive approach that prioritizes careful description, sensitivity to the institutions that produced the sources, and awareness of the historical and social context in which the individuals and families we study are embedded. Up-to-date information on each project is available at our group website.²

2 DATASETS

In this section we present the basic features and current status of each dataset. For this introduction, we organize our datasets into four categories: 1) family, kinship and demographic behaviour, 2) education, 3) employment, and 4) rural reconstruction. Each category includes a variety of datasets, some largely complete, and some still in progress. The populations are all Chinese, largely from the 18th, 19th, and 20th centuries. As of July 2020, the datasets include 8,167,457 records with nominative information on the behaviour and life outcomes of 1,753,700 individuals who were the focus of those records

1 See our group web page at <https://www.shss.ust.hk/lee-campbell-group/publications/> for a list of these award-winning publications and the datasets with which they are associated. Twelve of the authors are active or former group members: Cameron Campbell, Bijia Chen, Shuang Chen, Hao Dong, James Z. Lee, Lan Li, Chen Liang, Bamboo Y. Ren, Danching Ruan, Feng Wang, Emma Zang, and Hao Zhang. Other co-authors are largely from the Eurasia Project in Population and Family History notably Tommy Bengtsson, Marco Breschi, Satomi Kurosu, Matteo Manfredini, Michel Oris, Noriko Tsuya, and George Alter.

2 The Lee-Campbell Group website describes ongoing projects, affiliated faculty, students, and coders, and provides links to publicly released data and documentation: <https://www.shss.ust.hk/lee-campbell-group/>.

as well as several hundred thousand related individuals.³ Table 1 summarizes their contents. Three datasets and accompanying documentation are already available for download which we introduce in detail below with links and other information.

Table 1 *Lee-Campbell Group Datasets as of June 2020*

Name	Records	Persons
Family, Kinship and Population	2,975,211	488,675
Education	414,021	339,777
Employment	4,290,561	437,584
Rural Reconstruction	487,664	487,664
Total	8,167,457	1,753,700

What these diverse datasets have in common is that each seeks to record a substantively interesting well-defined target population in its entirety. This approach is different from constructing a statistically representative sample from a large target population in order to make inferences about it. By including entire target populations, we can describe broader social, political, and economic processes involving important or influential population subgroups in more detail than is possible with representative samples drawn from the general population. This is particularly important for upper-class populations such as nobles, civil servants, university faculty, and elite university graduates who may be important agents and/or harbingers of change, but account for only a small fraction of the general population and in a representative sample would include too few cases to allow much analysis of their composition, function, and change over time.

We introduce the datasets in detail below. The family, kinship and population datasets are defined by geographic area and hereditary status. Two are largely rural populations in northeast China, while the third consists of members of the Qing Imperial Lineage, almost all of whom lived in Beijing and what is now Shenyang. The education datasets include students from almost all the major Republican-era Chinese universities and from two major universities after 1949, as well as the vast majority of Chinese who graduated from foreign universities before the early 1950s. The employment datasets include nearly all civil and many military Chinese officials employed between 1760 and 1912, as well as separate nominative information for Chinese professionals from the fall of the Qing in 1912 to the early 1950s including almost all certified accountants, health professionals, engineers, and university faculty nationally as well as legal professionals in Shanghai and Beijing. Finally, the rural reconstruction datasets are drawn from nominative lists of individuals organized often by household for entire villages, rural brigades and communes, and even entire counties undergoing rural reconstruction either during Land Reform in the late 1940s or subsequent reorganization in the early and mid-1960s.

2.1 FAMILY, KINSHIP AND DEMOGRAPHIC BEHAVIOUR

Inspired by earlier studies on Chinese population history by Ping-ti Ho (1959) as well as by the contributions to English and French social economic history by Louis Henry (Gautier & Henry, 1958), Peter Laslett (Laslett & Wall, 1972) and their colleagues and students, our initial efforts at collecting individual level microdata for historical China focused on family, kinship and demographic behaviour. The datasets we eventually produced, collectively referred to as the China Multi-Generational Panel Datasets (hereafter CMGPD), are amenable to event-history analysis to examine community and family contextual influences on demographic behaviour and socioeconomic outcomes (Dong, Campbell, Kurosu, Yang, & Lee, 2015b). Table 2 summarizes the CMGPD datasets.

³ For the family, kinship and population datasets, each record only provided information about the person who was the focus of the record, akin to an entry in a census form. For many of the remaining datasets, in addition to details about the individual who was the focus, records sometimes also listed the names and sometimes occupation or other information of related individuals, typically their spouses, parents, or other relatives.

Table 2 *Family, Kinship and Demographic Behaviour Datasets*

Name	Acronym	Time Span	Households	Records	Persons	Started
China Multi-generational Panel Datasets	CMGPD					
Liaoning	CMGPD-LN	1774–1909	209,880 ^a	1,513,352	266,091	1982
Shuangcheng	CMGPD-SC	1866–1913	156,711	1,346,826	107,551	2004
Imperial Lineage	CMGPD-IL	1633–1933		115,033	115,033	1990
Total				2,975,211	488,675	

^a Refers to the period 1789–1913. The CMGPD-LN does not distinguish individuals by household before 1789, only by an administrative unit we refer to as household group.

The China Multigenerational Panel Database-Liaoning (hereafter CMGPD-LN) and China Multigenerational Panel Database-Shuangcheng (hereafter CMGPD-SC) are based on sets of population registers covering different administrative populations within Liaoning and Shuangcheng that in terms of format and organization resemble census-like listings of households and their members compiled at frequent intervals. Households and individuals are listed in roughly the same order in every edition, making the manual linkage of their records across time straightforward. Using the longitudinal data that we construct through record linkage, we can not only study the life histories of individuals, but also the histories of households and lineages. Because the content and organization of the CMGPD-SC and CMGPD-LN resemble those of population sources for other countries, they have been used in comparative studies, most notably in the Eurasia Project on Population and Family History introduced below.

The CMGPD-SC and CMGPD-LN both record household of residence, relationship to household head, demographic outcomes including birth, marriage, death, and basic measures of socioeconomic status. The populations they record are closed in the sense that entries and exits are rare, and when they occur, the timing is recorded, so at any given point in time the set of individuals at risk of experiencing an event is well-defined. Moreover, the target communities for each register series are recorded completely. Because the data are organized by residential household and are also multi-generational, and the detail on relationship to household head allows for children to be linked to their parents, grandparents, and other kin, they allow for us to embed individuals in their households and larger kin networks and examine how their life outcomes depend on the characteristics of distant kin and ancestors.

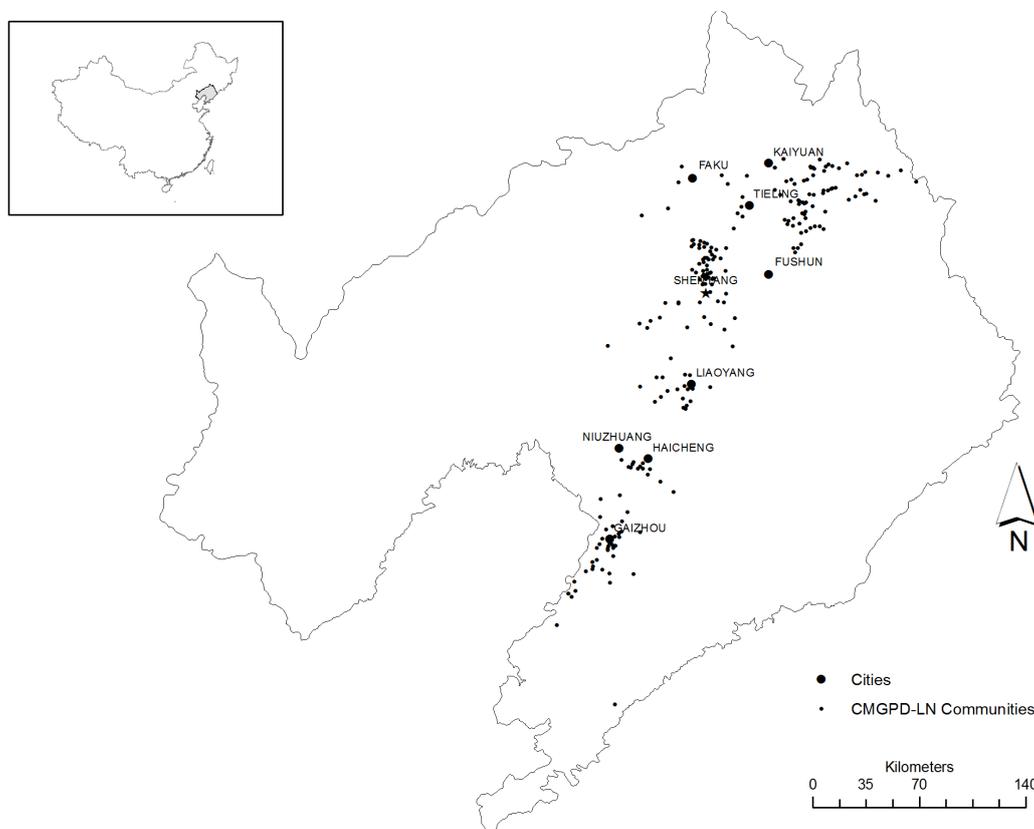
The China Multigenerational Panel Dataset-Liaoning includes 698 communities over a swath of what is now Liaoning province between 1749 and 1909 (Lee, Campbell, & Chen, 2010). The 2010 CMGPD-LN public release includes the contents of 732 population registers covering 29 administrative populations. The registers recorded each individual and household every three years. The resulting dataset includes 1.51 million records of individuals and 209,880 records of households. Through linkage of individual records, we reconstruct the histories of 266,091 individuals over as many as seven generations. We also link households over time. The data and accompanying documentation are available for download at the Inter-university Consortium for Political and Social Research (hereafter ICPSR).⁴ Each of these included a specific administrative population living in different geographic areas.

CMGPD-LN communities are scattered over a large area roughly equivalent in size to the Netherlands (see map 1), and were economically, ecologically, and geographically diverse, including coastal communities who relied on fishing as well as farming, inland communities who cultivated fruit orchards as well as dry field agriculture, and mountain communities which supplemented such activities with hunting and gathering (Ding, Guo, Lee, & Campbell, 2004). The series include populations of regular

4 For CMGPD-LN data and documentation, please visit <https://doi.org/10.3886/ICPSR27063.v10>. In the last three years, the documentation has been downloaded 8,377 times and data has been downloaded 2,821 times (According to <https://pcms.icpsr.umich.edu/pcms/reports/studies/27063/utilization> accessed on July 14, 2020).

farmers who were tenants on state land as well as specialized populations who supplied the state in-kind levies of fish, honey, mink pelts, and other goods. Unusually for an historical Chinese source, the registers record married and widowed women completely and in detail. Like many historical Chinese sources, however, they tend to omit children who died young, especially if they were female.

Map 1 *China Multigenerational Panel Dataset-Liaoning Communities*

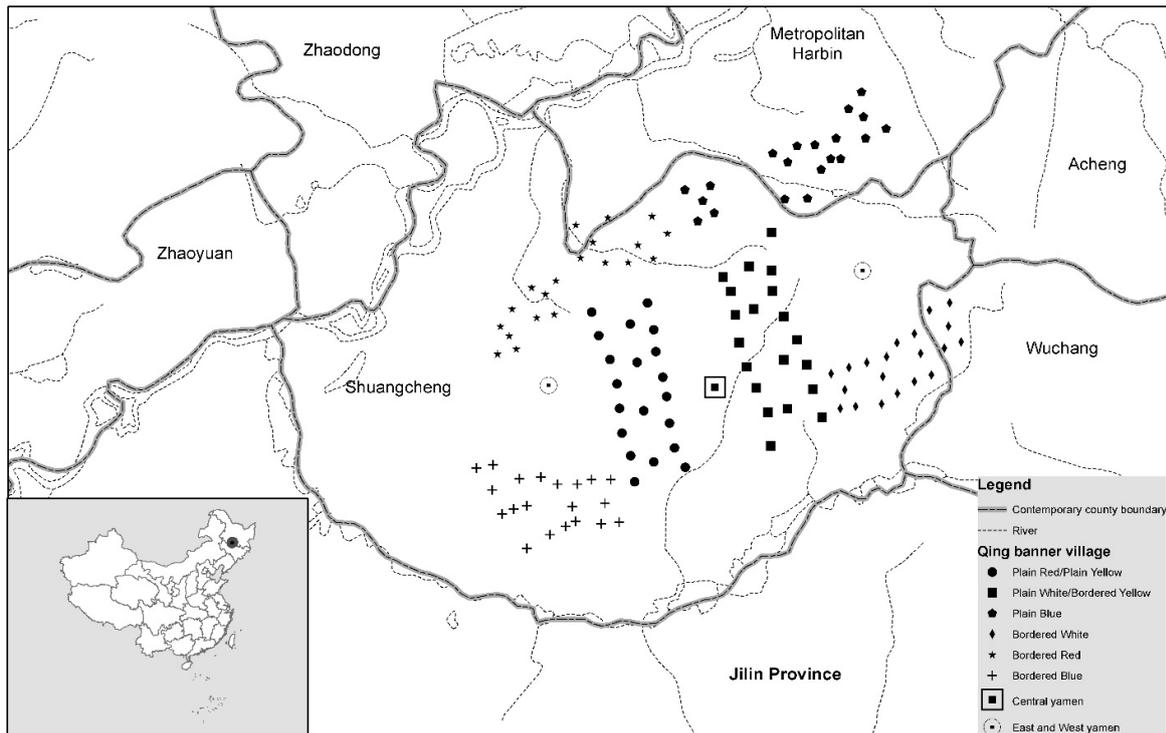


Source: Reproduced from page 4 of Lee, Campbell and Chen (2010).

The China Multigenerational Panel Dataset-Shuangcheng covers 129 communities in Shuangcheng county in Heilongjiang province from 1866 to 1913 (Chen, 2017; Wang et al., 2013). It contains 1,346,826 individual records and 156,711 household records. Through linkage, we reconstruct the histories of 107,551 individuals over as many as five generations. We also follow households over time. Map 2 presents the geographic locations of these communities and shows the location of Shuangcheng location in contemporary China. We publicly released the CMGPD-SC along with accompanying documentation at ICSPR in 2014.⁵ The data are annual and drawn from 14 separate register series. Unlike the CMGPD-LN which covers a much larger area, CMGPD-SC communities are confined to one 3000 square kilometre county directly south of the city of Harbin. Relative to the CMGPD-LN, the data in the CMGPD-SC have more detail on their socioeconomic characteristics. Reflecting the population's diverse origins, for example, the registers record each household's official ethnicity: Manchu, Mongol, Han, Xibo and others. Moreover, by linking the CMGPD-SC household records to separate CMGPD-SC land registers, the CMGPD provides data on each household's landed wealth distinguishing between assigned and acquired land. Like the CMGPD-LN, the CMGPD-SC records widows and married women in detail. It records somewhat more children who died young and more daughters than the CMGPD-LN, but such recording is hardly complete.

5 In the last three years the documentation has been downloaded 6,240 times and data has been downloaded 2,389 times (According to <https://pcms.icpsr.umich.edu/pcms/reports/studies/35292/utilization> accessed on July 14, 2020).

Map 2 Contemporary Shuangcheng County with CMGPD-SC Villages



Source: Reproduced from Wang et al. (2013, p. 5). Original by Matthew Noellert using historical geographic coordinates provided by Yuxue Ren, Shanghai Jiaotong University Department of History, and base map data from the Harvard Yenping Institute's China Historical Geographic Information System (2007).

The China Multigenerational Panel Dataset-Imperial Lineage (CMGPD-IL) records 115,033 members of the Qing Imperial Lineage and another 135,000 or so related individuals such as spouses from before the founding of the Qing dynasty in 1644 until 1933, two decades after the fall of the Qing (Cai, Lee, Campbell, & Myers, 1994; Lee, Campbell, & Wang, 1993; Lee & Guo (Eds.), 1994). 70,000 members are sons and daughters in the main line (*zongshi*) of the lineage and the remaining 45,000 or so are sons in the collateral line (*jueluo*) (Wang, Campbell, & Lee, 2010; Wang, 2012). In contrast with most lineage genealogies in China, these records were compiled prospectively by the Imperial Lineage Office (*zongrenfu*) which throughout the Qing employed 50 to 60 officials to record Imperial Lineage members, who resided almost exclusively in Beijing and Shenyang, and administer their affairs from cradle to grave. The 28 editions of the Jade Records (*yudie*) produced by the Office between 1660 and 1921 are among the most detailed and complete records of fertility and infant and child mortality for a large Chinese population before the mid-20th century. Unlike the CMGPD-LN or CMGPD-SC they do not record residential household composition. However, they do include almost all births, including daughters, official titles and employment, notable events, and the timing of exits via death and (for daughters) out-marriage (Ju, 1994). By contrast, the privately compiled lineage genealogies that were used in many studies of Chinese historical demography rarely record daughters or wives, and tend to omit sons who died in infancy, childhood, and even early adolescence, as well as adult males who never married or married but had not surviving sons (Campbell & Lee, 2002a). By the 19th century, the Imperial Lineage was internally diverse in terms of the social and economic status, including close relatives of the emperors who had a variety of privileges, and very distant relatives of the emperors whose status was mundane.

2.2 EDUCATION

As our research interests moved from population and family history to long-term trends in social mobility and social stratification, we expanded our collection to include individual student records in university archives.

We collectively refer to these datasets as the China University Student Datasets (hereafter CUSD). We distinguish between overseas students (OS) who graduated from foreign universities and domestic

university students during the Republic of China (ROC) and Peoples Republic of China (PRC). Based mostly on records of matriculating university students, these datasets typically include their name, major, place of origin, current address, previous education, and the names and occupations of their parents, and sometimes guarantors. The data accordingly not only provide information on students' family origins, but also for some students allow linkage of their records to those of their parents or siblings in the CUSD as well as to their and their family members' records in other datasets.⁶ Table 3 summarizes the CUSD.

Table 3 *Education Datasets*

Name	Acronym	Time Span	Records	Persons	Started
China University Student Datasets	CUSD				
Republic of China	CUSD-ROC	1912–1949	165,981	136,220	2010
People's Republic of China	CUSD-PRC	1949–2003	150,893	150,893	1998
Overseas Students	CUSD-OS	1847–1948			2015
In Japan			64,164	32,543	
In the United States			12,457	11,289	
In Europe			7,402	7,356	
In Soviet Union			766	758	
Other			87	81	
Not Specified			12,271	637	
Subtotal			97,147	52,664	
Total			414,021	339,777	

The China University Student Dataset-Republic of China (CUSD-ROC) covers university students in the Republican (1912–1949) era (Liang, Dong, Ren, & Lee, 2017; Ren, Liang, & Lee, 2020). It includes all or partial student registration records for 34 Republican Chinese universities. While these 34 universities represent only one-third of the universities during the Republic of China, they account for 90% of the surviving student registration records we have located in Chinese university and administrative archives. They include most of the major public, private, and missionary universities. As of January 2020, we have entered 165,981 records of 136,220 students from 34 universities. Almost all these records include student's major, age, gender, and place of origin. Most student records also include the names, occupations, and addresses of at least one parent, and in some cases of grandparents and guarantors as well. Entry of such information on parents, grandparents, and guarantors is still in progress, and we hope to add data from additional universities whose data we have already located.

The China University Student Dataset-People's Republic of China (CUSD-PRC) includes information for 64,500 undergraduate students who matriculated at Peking University between 1952 and 1999 and 86,393 undergraduate students who matriculated at Suzhou University between 1933 and 2003. The datasets are important both for their focus on elite university students in the People's Republic of China, and for their coverage of the last half of the 20th century. Peking University is one of the top national universities in China, and Suzhou is one of the best ranked regional universities. While censuses and retrospective surveys carried out since the 1980s identify college graduates, only a small number of recent surveys specify the university they attended, thus with only a few exceptions it was not previously possible to study the social or geographic origins of students at elite institutions, let alone from the 1950s to the present.⁷

⁶ Entry is ongoing and we expect the number of students linked to parents or other kin to rise substantially in the coming years, thus we do not present numbers here.

⁷ See pages 24–37 of Liang et al. (2013) for details on the construction of the CUSD-PRC, pages 37–46 for a discussion of the methods used in the analysis, and 46–57 for the contents. To protect privacy of the students in the records, entry was carried out on-site at both universities by university personnel and data remained there. Analysis of identifying data was also carried out onsite by university personnel. Analysis offsite relied on non-identifying tabulations, transformation or other calculations that had been conducted onsite.

The newest university student dataset, the CUSD-Overseas (OS), includes 52,664 Chinese students who pursued education overseas from the late 19th to the mid-20th century, accounting for 75 to 80% of the estimated 65,000–70,000 Chinese students who graduated from foreign universities during this period. As of June 2020, the dataset includes 64,164 records for 32,543 Chinese students in Japan, 12,457 records for 11,289 Chinese students in the USA, 7,402 records for 7,356 students in Europe, and over a thousand students who studied in the Soviet Union and elsewhere or for whom country of study is not available. While the CUSD-OS are based on Chinese and foreign government records of overseas Chinese students and graduates and rarely include information on students' family members, such information will be available via linkage for those students whose undergraduate records are in the CUSD-ROC.

2.3 EMPLOYMENT

We recently have turned to constructing a variety of large datasets on individual employment in the professions and in civil and military service in late imperial, Republican, and contemporary China. Table 4 lists our employment datasets.

Table 4 *Employment Datasets*

	Acronym	Time Span	Records	Persons	Started
China Government Employee Datasets	CGED				
Qing	CGED-Q	1760–1912	4,178,078	346,541	2013
Republic of China	CGED-ROC	1911–1949	51,806	30,365 ^a	2018
China Professional Occupation Datasets	CPOD				
Republic of China	CPOD-ROC	1911–1949	55,178	55,178	2016
People's Republic of China	CPOD-PRC	1949–1952	5,500	5,500 ^a	2015
Total			4,290,562	437,584	

^a Approximate because entry and cleaning are in an early stage and still ongoing.

The largest and most developed of these datasets is the China Government Employee Dataset-Qing (CGED-Q) (Chen, Campbell, Ren, & Lee, 2020; Ren, Chen, Hao, Campbell, & Lee, 2016). The core information for the CGED-Q comes from the *jinshenlu*, a roster of civil offices that was compiled every three months during the Qing and listed almost every regular, salaried civil office and included the holder's name, place of origin, banner affiliation (if any), location of post, job title, and other details. Positions ranged from high offices in the Six Ministries and other central government units down to low-level offices in county administrations. Each edition lists 13,000–15,000 employees.⁸ As of July 2020, we have entered 4,178,078 records of 346,541 officials for the period between 1760 and 1912.⁹ Most of these are from the period between 1830 and 1912, during which coverage of surviving editions is nearly complete. We are releasing the data in stages. Microdata for the period 1900–1912 are already available for download at the HKUST DataSpace and at a mirror site maintained by Renmin University Institute of Qing History.¹⁰

Nominative linkage of the records of the same official in successive editions in the CGED-Q allows us to construct and study their career histories. Linkage procedures depended on whether officials

8 The CGED-Q also includes some lists of military officials from a roster *zhongshubeilan* that originally was also compiled every three months. Each edition recorded 7000–8000 military officials.

9 Of these, 3,606,301 are records of civil offices and 518,596 are records of military offices. 284,264 officials started their careers as civil officials and 60,807 officials started their careers as military officials.

10 Data and documentation were first made available at the HKUST Dataspace in May 2019: <https://doi.org/10.14711/dataset/E9GKRS>. As of July 14, 2020, data has been downloaded 741 times at the HKUST Dataspace and the CGED-Q project page at the Lee-Campbell Group website has had 14099 visits. We are collaborating with the Renmin University Institute of Qing History for the release. They maintain a mirror site to facilitate access to the data and documentation for users in mainland China. As of May 7, 2020, the Renmin University mirror site has had 3471 unique visitors.

had a hereditary affiliation with the Eight Banners by virtue of their descent from the conquest elite who established the Qing in 1644. The officials who were not affiliated with the Eight Banners were mostly Han Chinese and can be linked based on their surname, name, and province and county of origin. The combination of these four attributes was almost always unique, and the primary challenge is addressing a relatively small number of cases where records of the same individual are not linked because their name is written slightly differently in two editions, usually because one character is replaced by another that looks or sounds similar. Linkage of officials who were affiliated with the Eight Banners is more difficult because instead of a province and county of origin, their banner affiliation was recorded. Moreover, most officials who were bannermen were Manchu or Mongol and did not have surnames recorded. The primary challenge is accordingly the opposite of the one we face for non-banner officials. While approximately 86% of the combinations of name and banner affiliation are unique, for the remainder who are not unique we use additional information to prevent records of different individuals who have the same name and banner affiliation from being linked together as if they were one person. We are also entering and linking information on the family backgrounds of officials and other characteristics from records of exam degree holders and other sources.

We have also begun to extend our coverage of employment to include government officials and educated professionals in the Republican era and the early years of the People's Republic of China. The data are important for our understanding of state building and the emergence of educated professionals as a distinct social group during this period. They also allow for comparisons of officials and some professionals who served in the Qing, Republican, and early Peoples Republic of China. For the CGED-ROC (Republic of China) we have entered 31,658 records of government officials who served between 1911 and 1949. These include with some overlap 9,988 officials from the Ministry of Education, the Ministry of Defence, the Academia Sinica, and the five administrative branches called the Control, Examination, Executive, Judicial, and Legislative *Yuan*, and 21,580 officials from the Transportation and Railroad Ministries. Common variables include name, sex, age, place of origin, education credentials, current employment and employment history. Acquisition of relevant materials continues, and this dataset should expand substantially.

The China Professional Occupation Datasets-Republic of China (CPOD-ROC) date back to 2016 when Bamboo Ren located related sources from the Liaoning Provincial Archive at FamilySearch (formerly the Genealogical Society of Utah). Bamboo subsequently worked with other group members, notably Yibei Wu, in archives and libraries at Beijing, Hangzhou, Nanjing, and Shanghai, to compile five discrete datasets. Of the 55,178 currently entered, 18% are medical doctors, 36% are university faculty, and another 36% are engineers. The remaining 10% are lawyers and certified accountants. Data entry is ongoing, and we expect the number of professions and especially the number of professionals for whom we have data to increase rapidly.

2.4 RURAL RECONSTRUCTION

Finally, motivated largely by student interests to better understand China's rural revolution, we have collected and continue to collect nominative individual level datasets, collectively referred to as the China Rural Reconstruction Datasets (CRRD), to research China's rural reconstruction especially during the third quarter of the 20th century. Table 5 summarizes the rural reconstruction datasets.

Table 5 *Rural Reconstruction Datasets*

	Acronym	Time Span	Households	Records	Persons	Started
China Rural Reconstruction Dataset	CRRD					
Siqing	CRRD-SQ	1946–1966	25,050	63,905 ^a	63,905	2013
Land Reform	CRRD-LR	1946–1948	82,514	423,759	423,759	2011
Total			107,564	487,664	487,664	

^a The household forms only record names and details of individuals who are over the age of 15 *sui*, that is 13.5 years Western years or older. Counts of the numbers of individuals below that age are provided on the original household form and were entered but no other details about them are available.

We do so because one of the defining features of 20th century China was the transformation of the world's largest agrarian society during this period. We are constructing two datasets that record information about individual and household experiences during the most dramatic stages of this process between 1946 and 1966, when the Chinese Communist Party carried out a nationwide redistribution of land and then gradually organized rural communities into agricultural cooperatives and ultimately People's Communes. The China Rural Reconstruction Dataset-Land Reform (CRRD-LR) was created to study the nationwide Land Reform Movement from 1946 to 1953. During this movement, local governments in many parts of rural China kept systematic records of land reform events and activities. These records include detailed individual- and household-level registers of property expropriation and reallocation and the political struggles that accompanied this redistribution of wealth. Currently the CRRD-LR contains county-wide data on the land reform experiences of over 80,000 households with approximately 400,000 individuals in Shuangcheng, Heilongjiang between 1946 and 1948.

The China Rural Reconstruction Dataset-Siqing (CRRD-SQ) is one of the most systematic and detailed sources available on social and economic change in rural China from before land reform in the 1940s up to the eve of the Cultural Revolution in 1966 (Xing, Campbell, Li, Noellert, & Lee, 2020). It is based on household social class registration forms compiled in rural areas around 1966 as part of the Socialist Education Campaign also known as the Four Clean-up (*Siqing*) Movement. The CRRD-SQ currently contains data from over 25,000 of these household forms, one quarter in collaboration with the Shanxi University Research Centre for Chinese Social History, from four provinces: Shanxi, Hebei, Inner Mongolia, and Guangdong. Each form records two to three pages of information per household, including their property holdings and occupations before and after land reform in the late 1940s, at the time when cooperatives were formed in the mid-1950s, and at the time of compilation in 1965 and 1966; the household head's social relations, a three-generation family history, and social, demographic, and political details on every household member over 15 *sui*, that is approximately 13.5 years or older according to Western ages.

3 HISTORY

Looking back over the last forty years, we can distinguish three distinct phases in terms of dataset construction and to some extent research. In the first phase, from 1979 to 1989, transcription and analysis were slow because of limitations in funding, technology, and support personnel. Work focused on demographic analysis of early iterations of the CMGPD-LN. In the second phase, from 1990 to 2010, data entry accelerated as stable funding became available to support a core team of full-time data coders, first in the USA and then in the People's Republic of China. Data coverage expanded to include the entire current CMGPD as well as the CUSD-PRC. In the third phase, from 2010 to the present, the range of population categories broadened to include government officials, professionals and other educated elites with the initiation of the CGED, the CPOD, and the CUSD-ROC and CUSD-OS, as well as such topics as rural reconstruction with the creation of the CRRD-LR and CRRD-SQ. Here we narrate progress across these three phases. The emphasis is on when and how each project were initiated, the participants and their contributions, and key transitions in terms of approach and scale.

3.1 PHASE 1 — THE BEGINNING, 1979–1989

We began more than forty years ago when James Lee began to look for quantizable nominative individual level microdata in historical archives in mainland China beginning with a winter-long visit to the First Historical Archives in Beijing in 1979. Inspired by the quantitative historians and social scientists who in the 1960s and 1970s transformed our understanding of family and population in past times in Europe and North America by the construction and analysis of datasets from archival sources, he hoped to do the same for historical China.¹¹

11 Funding during this decade came from a combination of internal Caltech resources as well as from external support from the National Endowment for the Humanities and from a Wang Fellowship in Chinese Studies. In addition, support from the Academia Sinica and the National Program for Advanced Study and Research in China funded nearly all our travel as well as some of our research expenses in Beijing, Shenyang and Taipei.

In 1982 at the advice of Deyuan Ju, Lee visited the Liaoning Provincial Archives and obtained microfilms of five household registers from Daoyi covering the period between 1774 and 1798.¹² Together with Robert Eng, an economic historian with prior experience with Japanese historical population registers, Lee developed a coding scheme and personally transcribed the contents of the 1774, 1780, 1786, and 1792 registers into a fixed-column format, first on paper forms and then into digital files.¹³ Lee also took a course in demographic methods in 1984 offered by the Graduate Group in Demography at the University of California, Berkeley. Working primarily with various California Institute of Technology (Caltech) undergraduates, he published the first analyses of Chinese mortality, fertility, and household structure for specific historical populations on the Chinese mainland before the 20th century using household register microdata (Lee, Anthony, & Suen, 1988; Lee & Eng, 1984; Lee & Gjerde, 1986).¹⁴ He also published with William Lavelly and Feng Wang an influential article on how new historical and contemporary microdata were reshaping the understanding of Chinese demographic behaviour (Lavelly, Lee, & Wang, 1990).

While Lee acquired additional 19th century household and population registers for Daoyi beginning in 1985, research using such microdata did not advance significantly until the arrival in the summer of 1987 of Cameron Campbell, a Caltech sophomore (second year student) majoring in Electrical Engineering with a side interest in Chinese history he had developed in high school.¹⁵ Campbell had prior training and experience with database programming. After going over the various C programs that had been written for specific data transformations and calculations for the Daoyi registers, he proposed to Lee a new workflow where the data would be managed in dBase III+ (later dBase IV) and then exported to SPSS for analysis. Campbell began to develop the new code in summer 1987 when he and Lee re-visited the First Historical Archives in Beijing and the Liaoning Provincial Archives in Shenyang.¹⁶ Processing included construction of flag variables for the occurrence of demographic events, identifiers for records of the same individual in different registers and links between kin, measures of household structure and composition, and measures of context at the individual level including the presence and absence of specific kin. This simplified data entry and created new possibilities for analysis to move beyond the calculation of rates and proportions.¹⁷

3.2 PHASE 2 — ACCELERATION, 1990–2009

Entry of the CMGPD-LN subsequently accelerated, first in 1990 thanks to support from the Academia Sinica and the National Science Council in Taiwan, and again in 1999 thanks to a large private gift to James Lee. Before 1990, only one of the 29 administrative populations that eventually made up the CMGPD-LN, Daoyi, had been entered, along with a few registers from another administrative population, Gaizhou, yielding about 100,000 observations. In contrast, between 1990 to 1999, we entered an additional 400,000 records from 8 more administrative populations. Our increased speed

12 Additional registers were filmed at the Liaoning Provincial Archives in 1985 and 1987 that extended the series to 1873.

13 See Lee and Campbell (1997, pp. xix–xxi) for complete lists of the people who helped with the entry and analysis of the Daoyi registers, the coders who entered the data, and the funding sources.

14 By this time, Arthur Wolf had already done related work for Taiwan using Japanese colonial records from the first half of the 20th century (Wolf & Huang, 1980).

15 Campbell's participation in summer 1987 and 1988 was supported by Caltech Summer Undergraduate Research Fellowships.

16 To carry out programming and analysis while in Beijing and Shenyang in summer 1987, Lee and Campbell brought a large, very heavy portable Compaq microcomputer with them from the United States. We however forgot that voltage transformers were not widely available in China and therefore in Shenyang had to rely on one that had been hand-built and which was turned on and off with a knife-switch.

17 By the time Campbell graduated in June 1989, he had switched from Electrical Engineering to a double major in History and Engineering and Applied Science. After graduation, he went to Taipei and Beijing with funding from the Watson and Durfee Foundations to learn Chinese and then spend time in Daoyi. In Taipei he spent six months studying Chinese intensively at the Mandarin Training Centre at National Taiwan Normal University. He spent the summer of 1990 working with James Lee in Taipei. In fall 1990, he went to the University of Pennsylvania to study for a PhD in demography. He spent another year in Taipei for further intensive training in Chinese in 1992–1993 at the Inter-University Program in Chinese Language Studies at National Taiwan University and then spent summer 1993 conducting dissertation research in Beijing. He was supported by a Social Science Research Council International Pre-Dissertation Research Fellowship.

was partly due to the acquisition of most of the historical household registers and related materials from the Liaoning Provincial Archives by the Genealogical Society of Utah who made these sources available to us, and partly because the increase in funding enabled us to support a larger team of coders to transcribe these data into digital datasets.¹⁸ Almost all the data entry was done by coders in the United States, though two CMGPD-LN series were completed in Taiwan. In 1999, we moved data entry to mainland China where we were fortunate to have the support of one and soon afterwards three reliable and enthusiastic full-time coders: Xing Xiao, Huicheng Sun, and Jiyang.¹⁹ Over the next four years they entered the remaining 19 CMGPD-LN datasets, yielding an additional 1 million observations, and subsequently spent a full person-year largely in 2010 to clean the entire 1.5 million observation CMGPD-LN in preparation for public release of these data.²⁰

To conduct more detailed analyses of fertility and infant and child mortality than was possible with the CMGPD-LN, Lee collaborated with Huimin Lai and Sufen Liu at the Academia Sinica in Taiwan to construct the CMGPD-IL beginning in 1990. Deyuan Ju at the First Historical Archives had earlier introduced Lee to the collections of nominative historical demographic microdata from the Office of the Imperial Lineage including the Jade Registers (*yudie*) or Imperial Lineage Genealogy (Ju, 1994). Recognizing that the nearly complete recording of male and female births and infant and child deaths made the *yudie* an invaluable complement to the CMGPD-LN, which recorded few daughters, and omitted some sons who died in infancy or childhood, Lee obtained copies of these data in 1985 when the First Historical Archives in Beijing microfilmed the *yudie* for the Genealogical Society of Utah. Lee oversaw the coding of the members of the Main Line (*zongshi*) in collaboration with Huimin Lai and Sufen Liu between 1990 and 1992. Lee also recruited Feng Wang in 1989 to participate in the analysis of these data and published with Campbell an introduction to the CMGPD-IL dataset (Lee, Campbell, & Wang, 1993).²¹ Campbell used the CMGPD-IL in his PhD dissertation which studied long-term trends in mortality in Beijing by comparing mortality patterns in the CMGPD-IL in the 18th and 19th century with mortality patterns in Beijing in the 1920s and 1930s and after 1949.²² Later, working with Linlan Wang, a sociology PhD student of Lee's at Peking University, Lee added the sons from the collateral line (*jueluo*) recorded in the 1933 *Aixin Jueluo Genealogy* for her 2012 PhD thesis (Wang, 2012; Wang, Lee, & Campbell, 2010).²³

In 2003, Lee who was now at the University of Michigan, began to work with Shuang Chen, a PhD student in history, to construct the CMGPD-SC to examine the relationship between landholding and demographic behaviour in Shuangcheng, Heilongjiang (Chen, 2009; Wang et al., 2013).²⁴ In addition to landholding, the CMGPD-SC allows for other analyses not possible with the CMGPD-LN, for example comparisons by registered household ethnicity. The Genealogical Society of Utah had acquired 338 population registers and 23 land registers for Shuangcheng and in fall 2003 made them available to us. Xing Xiao, Huicheng Sun, and Jiyang entered the 1.3 million records in the CMGPD-

18 Upon completion of the CMGPD-LN, we provided a name list linking records to their pages in the original sources to the Genealogical Society of Utah (now [FamilySearch](#)) for inclusion in their online search facility.

19 Another logistical challenge after moving data entry to China was processing payments to the coders. We are grateful to staff at University of Michigan ICPSR, UCLA CCPR and HKUST for their oversight and management of the grants, especially Ruth Danner at ICPSR, Lucy Shao at UCLA CCPR, and Freda Ching at HKUST.

20 Between 1999 and 2006, we also visited some 57 Liaoning villages during which we acquired almost 250 related data sets such as family genealogies.

21 Lee, Campbell and Wang (1993, p. 361) lists the scholars at the Academia Sinica who facilitated this work, the coders who entered the data, the Caltech undergraduates who helped with programming, and the funding sources. Lee and Guo (Eds.) (1994) provides a detailed history in Chinese of the project, including the data entry process and the transformation of the data to prepare it for analysis.

22 Campbell collected the mortality data for 20th century Beijing while doing archival and library research there in summer 1993. While conducting this research he was assisted by Jennifer Huang Bouey.

23 Unlike the Imperial Lineage Genealogy but like almost all other historical Chinese genealogies, the Aixin Jueluo Genealogy did not report daughters. Data entry was done principally by Jiyang with the assistance of Xing Xiao from April 2008 to July 2011.

24 Wang et al. (2013, p. viii) provides a complete list of the individuals who contributed in various ways to the construction of the CMGPD-SC.

SC between 2004 and 2007.²⁵ By contrast, entry of the 1.5 million records in the CMGPD-LN had taken approximately twenty years. Shuang Chen, who is now an associate professor of history at the University of Iowa, oversaw the data coding and led the analysis of the CMGPD-SC for her dissertation (Chen, 2009) and book (Chen, 2017) as well as for our contributions to the Eurasia Project fertility volume (Chen, Campbell, & Lee, 2014).

Our capacity to manage and analyse steadily larger datasets improved as a result of additional changes in the workflow. The dBase programs that read in the files provided by the coders to produce the files to be analysed in SPSS and later STATA were maintained by Chris Myers in the early 1990s and then again by Cameron Campbell. However, the programs were slow. In the late 1980s, when there were fewer than 70,000 records to deal with, the dBase programs that transformed the raw data entered by coders into files to be used in analysis could run for more than a day and were prone to crashing.²⁶ Subsequent improvements in processing and disk speed were offset by increases in the number of records to be handled. Finally, in the mid-1990s, we froze development of the dBase programs. They continued to be used to process incoming files and prepare a file for analysis in STATA, but they were not further developed. Creation of new variables was done in STATA. Eventually, Campbell retired the dBase programs completely and wrote STATA code to handle the entire process of importing files, organizing them, creating variables for analysis, and carrying out analysis. This reduced the time required to go from the raw files provided by coders to the work files used in analysis to just a few hours.

To better understand the context of the communities that were recorded in the CMGPD-LN and learn more about the histories of the families it recorded, we conducted fieldwork in rural Liaoning. Between 1999 and 2006, we made eight field trips to Liaoning accompanied by Gao Jing and colleagues from the Liaoning Provincial Gazetteer Office and local Gazetteer Offices during which we visited 57 largely rural communities. We spent some 250 person-days visiting descendants of the CMGPD-LN populations and collecting local sources such as genealogies, tomb inscriptions, deeds, and other family documents on these populations. We also gathered oral histories and collected information on the families during the time from the end of the CMGPD-LN in 1911 up to the time of our visit. We compared these local data with the state household and population registers in Campbell and Lee (2002a) and Ding, Guo, Lee, and Campbell (2004). In each community, we shared with the families we visited genealogies of their lineages generated from the CMGPD-LN. Many families had lost their genealogies or had only rudimentary genealogies that listed only the generation and given names of male lineage members, and the materials we provided helped them reconstruct their family histories, including names and other information on ancestors who held office or had other achievements or recognition.

Taking advantage of advances in technology, we began to use event-history analysis and other regression-based approaches to study associations between individual demographic behaviour and outcomes and household and community context. The capacity of the personal computers that we were using for processing and analysis improved dramatically. In the early 1990s, computations involving the 100,000 or so records in the Daoyi series that involved anything more than tabulation or linear regression took fifteen minutes to an hour, depending on the number of observations included, the number of variables, the type of model, and the number of models. By the late 1990s, more advanced estimations run on much larger numbers of records took much less time. By the late 2000s, calculations involving the combined CMGPD-LN and CMGPD-SC files, nearly 3 million records, could be completed in minutes on a personal computer.

Our shift to event-history analysis was also spurred by Akira Hayami's 1993 invitation to James Lee to participate in the Eurasia Project in Population and Family History, an international comparative project that studied interactions between community context, household organization, and demographic

25 During this time, our work received support from NIH NICHD 1R01HD045695-01A2 (Demographic Responses to Community and Family Context, James Lee PI).

26 While Campbell was still at Caltech, the programs were so slow and prone to crashing that when he and Lee processed the dataset to integrate newly coded registers or reprocessed the dataset to add new variables, he often slept on the floor in Lee's office, waking up every few hours to check that the programs were still running, and if they had failed, correcting the problems and restarting them.

behaviour in past times.²⁷ The project began in 1994 and yielded three volumes in a dedicated series published by MIT Press under the editorship of Lee, Bengtsson, and Alter. The first was on mortality (Bengtsson, Campbell, Lee et al., 2004), the second was on fertility (Tsuya, Wang, Alter, Lee et al., 2010), and the third was on marriage (Lundh, Kurosu et al., 2014).²⁸ Teams of researchers who had household register data from communities in Belgium, China, Italy, Japan, and Sweden collaborated to specify event-history models of mortality, fertility and marriage that could be estimated in all of the datasets which would yield results that could be compared. We discuss findings from the project below.

In the early 2000s we began to plan for the public release of the CMGPD. Lee's affiliation with the Inter-university Consortium for Political and Social Research (ICPSR) after his move to Michigan in 2003 was crucial. Conversations with Myron Guttmann and others convinced us that public release of the CMGPD was not only important but, with support from ICPSR, also entirely feasible. ICPSR provided administrative support for the submission of funding applications to the National Institutes of Health and management of the resulting grants, and staff support for the deposit of CMGPD data and documentation through the ICPSR Data Sharing for Demographic Research (DSDR) program.²⁹ After James Lee moved to the Hong Kong University of Science and Technology (HKUST) in 2009, Cameron Campbell at University of California, Los Angeles (UCLA) received administrative support from the California Centre for Population Research for submission and management of a grant to support public release of the CMGPD-SC, but the release itself was again through ICPSR DSDR with support from ICPSR staff, most notably Susan Hautaniemi Leonard. From 2011 to 2014, Campbell conducted training workshops at Shanghai Jiao Tong University every summer to introduce prospective users to its contents and organization and demonstrate advanced operations to manage and analyse the data. Yuxue Ren hosted him and Dan Xu provided logistical support. Based on experiences from the workshops, Campbell, Hao Dong and James Lee produced the *CMGPD Training Guide* (Campbell, Dong, & Lee, 2013) as a companion to the *User Guides* (Lee, Campbell, & Chen, 2010; Wang et al., 2013).

In the late 2000s, we initiated a new line of work on social mobility, stratification, and inequality. We had originally considered father-son associations in social and economic outcomes in Daoyi in Lee and Campbell (1997, pp. 196–214). When Campbell first started as an assistant professor in sociology at (UCLA) in 1996, he was still focused on demography, but interactions with Donald Treiman, William Mason, Robert Mare, Judith Seltzer, Ken Sokoloff, Jean-Laurent Rosenthal and other colleagues in sociology and economics inspired a desire to take advantage of the distinctive properties of the

27 See the acknowledgments in Bengtsson, Campbell, Lee et al. (2004, pp. xi–xii), Lundh, Kurosu et al. (2014, pp. xxiii–xxv), Tsuya, Wang, Alter, Lee et al. (2010, pp. xxiii–xxv) for a history of each volume and Lundh, Kurosu et al. (2014, pp. xvii–xxi) for a recap and reflection. Campbell (2016) is a personal reminisce and reflection on his involvement in the project. See Lee and Steckel (2006) for an appreciation and appraisal of Bengtsson et al. (2004). Daniel Little has discussed the Eurasia Project in Population and Family History in a post on his blog *Understanding Society*: <https://understandingsociety.blogspot.com/2014/08/eurasia-project-on-population-and.html>.

28 We would also like to thank the many participants who discussed these books at authors-meet-critics sessions at the Social Science History Association in 2004, 2010, and 2015 and the Population Association of America meetings in 2005 and 2011, and at the summary discussion of the Eurasia Project in Population and Family History at the 2014 American Sociological Association. Among the many participants at these sessions were Douglas Anderton, Jason Beckfield, Hilde Bras, Andrew Cherlin, Jack Goldstone, David Hacker, Michael Haines, Charles Hirschman, Jan Kok, Ronald Lee, Daniel Little, Deirdre McClosky, Myron Guttmann, Karen Oppenheim Mason, Richard L. Steckel, Jan Van Bavel, and Andreas Wimmer. Before the publication of each volume, work in progress was presented at dedicated sessions at Social Science History Association, European Social Science History Conference, World Economic History Congress, and other meetings. These sessions are listed in the acknowledgments of the respective volumes.

29 NIH NICHD 1R01HD057175-01A (The Liaoning Multi-Generational Panel Dataset: Public Release and User Training, Lee and then Leonard PI) supported release of the CMGPD-LN between 2009 and 2012. NIH NICHD 1R01HD070985-01 (Multi-generational Demographic and Landholding Data: CMGPD-SC Public Release. Campbell PI) supported public release of the CMGPD-SC between 2012 and 2016.

CMGPD-LN to study social mobility, inequality, kinship and other topics.³⁰ We moved past father-son associations to study associations of socioeconomic outcomes with the characteristics of progressively wider networks of kin, starting with siblings, uncles and grandparents in the same household, then moving to kin outside the household, and then eventually to lineages. We faced a constraint, however, in that the CMGPD datasets only recorded official positions held by adult males. They provide no information about other non-agricultural occupations, and only the CMGPD-SC recorded landholding.

Inspired by this growing interest in inequality and social mobility and a desire to move beyond the study of demography and after having devoted over two decades to the collection of individual level information on socio-economic attainment and related demographic behaviour in pre-20th century China, we turned our attention to the construction of datasets for the study of inequality, social mobility, and social change in historical and contemporary China. Lee and collaborators initiated the CUSD-PRC when Lee learned of the student registration cards held in the Peking University archives. Actual entry of data at Peking University began in 2003 under the supervision of Lee's collaborators, Danching Ruan and Shanhua Yang, and Yang's PhD student Hao Zhang. Entry of data at Suzhou University began in 2006, supervised by Chen Liang, a recent Tsinghua University history PhD, who went on to a postdoc at Michigan mentored by James Lee, and is now a history professor at Nanjing University.

3.3 PHASE 3 — THE EXPANSION, 2009–PRESENT

The most recent phase began when James Lee moved to the Hong Kong University of Science and Technology in 2009. Campbell moved from UCLA and joined him in 2013. Proximity to the mainland allowed for frequent travel there to interact with researchers, present results, and explore new sources of systematic information on China's historical and recent past. This led to the discovery of new materials that could be transformed into databases and which allowed us to move from the study of household organization and demographic behaviour to the study of educated elites. Generous and stable intramural and extramural funding supported a further expansion of coding efforts.³¹ It was also easier to recruit, support, and work with graduate students with relevant interests than was the case while Lee and Campbell were in the United States. Since Lee's arrival at HKUST, we have trained and collaborated with Matt Noellert, Hao Dong, and Bijia Chen who are now respectively tenured faculty in economic history at Hitotsubashi University, tenure-track faculty in sociology at Peking University, and postdoctoral fellow in history at Renmin University's Department of History. As of the summer of 2020, our current PhD students are Xiangning Li and Bamboo Y. Ren at the Hong Kong University of Science and Technology, Qin Xue at Central Normal University, and Li Yang and Yibei Wu at Shanghai Jiaotong University.

We now construct datasets for the study of inequality, social mobility, and social change in historical and contemporary China. In 2010, Chen Liang proposed a study of social origins of university students in the first half of the 20th century that would extend on the CUSD-PRC by construction and analysis of a dataset (CUSD-ROC) based on student registration cards held in historical archives throughout China (Liang et al., 2017; Ren et al., 2020). Working with others in the Lee-Campbell Group, he located the student registration cards for half of the 34 universities that currently make up the CUSD-ROC and arranged for most of their data entry. Bamboo Y. Ren, James Lee, and Mingyu Zhang located and coded the other half of these data. The process for transcription of these records differed from that of the CMGPD datasets. Rather than having a dedicated team transcribe the contents of scans of the original sources, data was entered on-site in the archives by personnel recruited locally for the task. Coding additional variables or checking the originals to resolve inconsistencies required return trips to the archives.

The next project we initiated was the CGED-Q. Cameron Campbell conceived of it as a resource for the study of Qing officialdom and the career dynamics of Qing officials in 2013 when Yuxue Ren, a postdoctoral fellow of Lee's at the University of Michigan and now an associate professor of history

30 In the late 1990s and through the 2000s, three seminar series at UCLA were especially important in inspiring Campbell to move beyond traditional historical demography: the Von Greppe Workshop in Economic History, the UCLA/RAND Joint Labour and Population Workshop, and the California Centre for Population Research Seminar Series.

31 Our work has been supported by a series of grants from the Hong Kong Research Grants Council General Research Fund: 642911 (Lee PI), 640613 (Lee PI), 16400114 (Campbell PI), 16400714 (Lee PI), 16602315 (Lee PI), 16600017 (Campbell PI), 16602117 (Lee PI), 16601718 (Campbell PI).

at Shanghai Jiao Tong University, showed him and James Lee work she was doing with records of officials in northeast China she had transcribed from quarterly *jinshenlu* in a collection of 206 editions published by the Tsinghua University Library.³² Campbell, Lee, and Ren developed a plan to enter all of the 2.8 million records in this collection and an additional 1.2 million records from *jinshenlu* editions held elsewhere. This was completed in summer 2020. The coders we had relied on for the CMGPD began entering data in 2014. In 2016 we added new coders and the pace of entry doubled. Bijia Chen joined the project in 2015 while she was an MPhil student in Social Science at HKUST. She played a key role in the coordination of the data entry and then wrote her PhD dissertation on the careers of Qing officials (Chen, 2019).

While the CMGPD, CGED, and much of the CUSD were all initiated either by James Lee or Cameron Campbell, working with other senior members of our research team such as Chen Liang, our most recent data sets, on China's rural revolution (CRRD-LR and CRRD-SQ) and the rise of China's professionals (CPOD) are the initiative largely of our younger team members who found these sources and organized the data construction for their PhD theses. Matthew Noellert discovered the materials that became the basis of the CRRD-LR while conducting fieldwork in Shuangcheng in 2011 and used these data to write his 2014 PhD dissertation and 2020 book.³³ Moreover, while Long Xing, the Director of the Shanxi University Research Centre for Chinese Social History, initiated the China Rural Reconstruction Dataset-Siqing (CRRD-SQ) based on his collection of 7800 rural household social class registration forms, it was Noellert, working together with Yingze Hu, who oversaw the initial coding of the CRRD-SQ in 2015³⁴, and Noellert together with HKUST PhD student, Xiangning Li, who between 2016-2019 expanded the CRRD-SQ from one to four provinces, and from eight thousand to 25 thousand households.

Similarly while the CUSD-ROC and CUSD-OS were initiated by Chen Liang in 2010 and James Lee in 2019, it was Bamboo Ren who initiated and coordinated the construction of the various CPOD datasets beginning in 2018 and Yibei Wu and Li Yang working under Bamboo's direction who coordinated the data entry of the CUSD-OS, and Yibei Wu who located the data sources and oversaw the data transcription of the CGED-ROC.

4 FINDINGS

We organize our review of findings by topic. We begin with studies of demographic behaviour and household organization. We follow the development of these studies from estimates of demographic rates and household structure to examinations of the implications of family hierarchy based on patterns of differentials according to household context and finally to analyses of assortative mating, relationships between household context and health and mortality in later life, and other topics. We then present work on intergenerational social mobility and inequality more generally. We start with the earliest studies of the associations between fathers' and sons' socioeconomic outcomes, move to multi-generational studies that considered the role of kin other than the father in shaping individual outcomes, and conclude with recent studies that move beyond the individual to consider kinship networks and descent groups as units of analysis, and stratification and inequality more generally. Third, we summarize recent published work on the geographic and social origins of university students in 20th century China. Finally, we summarize recent published studies of the careers of government officials during the Qing.

32 See Chen et al. (2020) for an English language introduction and Ren et al. (2016) for a Chinese language introduction to the project, sample of results, and complete list of the collaborators, students, and coders who contributed to the CGED-Q.

33 Huicheng Sun, Xing Xiao, and Jiyang coded the CRRD-LR largely during 2012 and linked individuals across various CLRD event registers in 2013.

34 The resulting dataset is held at Shanxi University and analysis is carried out by personnel there. As with the CUSD-PRC, any analysis we conduct is based on tabulations or other calculations produced in response to our requests, not on the original data. Since 2016, we have located and entered another 18,000 forms. Matt Noellert and Xiangning Li lead the analysis of these materials.

4.1 DEMOGRAPHIC BEHAVIOUR

The earliest line of work described trends and patterns in mortality, fertility, population age and sex composition, and household structure by presentation of aggregated rates and proportions. Based on five registers from Daoyi between 1774 and 1798, Lee and Eng (1984) introduced the data and presented descriptive results on birth and death rates, population age composition, and household structure. Among other findings, they showed that these sources recorded adult males and married and widowed females completely but omitted many sons who died in infancy or early childhood along with most daughters. Lee, Campbell, and Wang (1993) introduced the CMGPD-IL and presented time trends and age patterns of mortality for the members of the Imperial Lineage. Lee, Anthony, and Suen (1988) and Lee, Campbell, and Anthony (1995) showed that levels and patterns of mortality in Daoyi resembled those in other historical populations. Lee and Gjerde (1986) compared household forms in Daoyi with those in Norway and the United States to show that existing schemes for classifying household structure were inadequate and proposed a new classification scheme more amenable to comparison between Europe and other societies. Comparison to the CMGPD-LN revealed that adult Imperial Lineage males had higher death rates than adult males in Daoyi in rural Liaoning, presumably because being restricted to living in Beijing subjected them to an 'urban penalty'. Results from these early studies led to lines of research exploring relationships of fertility, mortality, and other aspects of demographic behaviour to other social and economic variables which took advantage of the individual-level detail in the data. We describe these lines of research below.

4.1.1 FERTILITY

Early results on fertility patterns in these studies led to a line of work on the role of deliberate delay or cessation of childbearing in producing low levels of fertility within marriage. Wang, Lee, and Campbell (1995) and Lee and Campbell (1997, pp. 83–102) showed that in the CMGPD-IL and Daoyi respectively, marital fertility was lower than in Europe, intervals between marriage and first birth and between subsequent births were much longer, and childbearing ceased much earlier. They argued that these and other patterns were consistent with deliberate behaviour to delay births and cease childbearing. These findings were the basis of the claim in Lee and Wang (1999) that contrary to the beliefs of Malthus and his successors, a fertility-based preventive check played an important role in the population dynamics of China before the 20th century, and that the rapidity of fertility decline in the 20th century in mainland China, Taiwan and Hong Kong reflected a historical legacy of adjusting fertility according to economic and other circumstances that primed the population to respond quickly when new technologies for fertility limitation became available. A vigorous debate with advocates of a Malthusian interpretation of China's historical population dynamics ensued (Campbell, Wang, & Lee, 2002; Lee, Campbell, & Wang, 2002). Campbell and Lee (2010b) revisited the issue of fertility control and showed that once heterogeneity in fecundity across couples was properly accounted for, there was clear evidence of stopping behaviour.

Paralleling our study of mortality, we moved on to map fertility differentials to illuminate influences of community, household and individual context on reproduction. Lee and Campbell (1997, pp. 133–156, pp. 177–195) compared cumulative numbers of boys born by household structure, location within the household, and socioeconomic status in Daoyi. In general, men who had privileged statuses in the household or socioeconomic hierarchy had more children. This reflected not only earlier marriage and a higher likelihood of remarriage, but in some cases higher fertility within marriage. This consistent positive relationship between privilege and reproduction is in contrast with mortality, which as noted above was in some cases the opposite of what was expected, with privileged males experiencing higher death rates. Fertility fell during times of hardship, that is when grain prices were high or when there were climatic shocks (Campbell & Lee, 2010a; Wang, Campbell, & Lee, 2010). Dong (2016) also studied the role of local family system in moderating the influence of co-resident kin on reproduction between East Asian populations.

Subsequent analyses focused primarily on fertility within marriage and such related behaviour as adoption. Wang, Lee, and Campbell (2010) revisited fertility in an expanded CMGPD-LN and demonstrated that it was linked to location in economic and household hierarchies. Campbell and Lee (2009) examined associations of marital fertility with characteristics of kin living outside the household but did not find any. Chen, Lee, and Campbell (2010) showed that fertility in Shuangcheng was positively associated with family landholding and with other measures of socioeconomic and household status. Wang and Lee (1998) showed that in the Qing imperial lineage, as many as 12.5% of sons

were adopted between related individuals, and that more generally, adoption played an important role in maintaining the continuity of the descent line and achieving other goals.

4.1.2 MORTALITY

Early descriptive analysis of infant and child mortality led to an examination of female infanticide that became one of the foundations of the critique of Malthusian interpretations of China's historical population dynamics in Lee and Wang (1999). Lee, Campbell, and Tan (1992) and Lee and Campbell (1997, pp. 58–82) used indirect evidence from registered births and deaths in Daoyi to argue that families employed female infanticide or neglect to influence the number and sex composition of surviving children, and in doing so responded to economic conditions as well as their personal circumstances. As noted above, dissatisfaction with reliance on indirect evidence inspired the creation of the CMGPD-IL, which recorded the births of sons and daughters completely, as well as the deaths that occurred afterward. This led to the analysis of infant and child mortality in the Imperial Lineage in Lee, Wang, and Campbell (1994) that provided direct evidence of infanticide in the form of dramatically higher death rates for daughters in the first day and month of life, furthermore demonstrating that infanticide was not only a response to crisis or desperate poverty.

The next set of mortality studies employed event-history analysis to map patterns of mortality differentials and illuminate the influence of family, community, and institutional context on death risks. Lee and Campbell (1997, pp. 133–156, pp. 177–195) first showed that mortality rates varied according to socioeconomic status and location within the household hierarchy. Relationships were sometimes counterintuitive: male privilege was sometimes associated with higher mortality risk. Campbell and Lee (1996) showed that mortality risks depended not only on household size and composition but also on the presence or absence in the household of specific kin. Campbell and Lee (2002b) examined how household context conditioned the mortality effects of widowhood and orphanhood and showed that widows' mortality risks depended on whether they had a son. Widows who had a son were unaffected by the loss of their husband, but widows without a son experienced elevated mortality. As we describe below in our summary of findings, Hao Dong lead-authored comparisons of family contextual influences on mortality in Liaoning and Taiwan in China and northeast Japan using pooled household register data from all three locations (Dong, 2016; Dong et al., 2017).

This led to studies of the short-term consequences of economic and climatic shocks and long-term effects of public health interventions. Campbell and Lee (2000) argued based on an analysis of effects of the interactions of social status, household context and prices that there was a trade-off between privilege and mortality risk, with the mortality of privileged individuals also being more sensitive to price fluctuations. Campbell and Lee (2004) used a much larger sample from the CMGPD-LN to investigate differentials in mortality levels and mortality sensitivity to price fluctuations in more detail. Male mortality was more sensitive to grain price fluctuations than female mortality, and the response was conditioned by age, socioeconomic status, and household context. These results contributed to the comparisons between East and West in Bengtsson et al. (2004). Campbell and Lee (2010a) investigated the effects of unusually cold summers and other climatic disruptions in the years 1782–1789, 1813–1815, and 1831–1841. During the first of these periods, life expectancy fell by more than 10 years. Young males and females were especially hard hit, with the death rates of males aged 5–15 being multiplied by 8.78 and females aged 5–15 being multiplied by 4.65. Campbell (1997; 2001) assessed the effects of public health interventions in Beijing at the beginning of the 20th century and immediately after 1949 by comparing mortality in the CMGPD-IL in the 19th century to rates in Beijing at different points in time in the early, mid and late 20th century.

Recent studies investigate the consequences of family context and history for mortality later in life or in later generations. Chen et al. (2005) used the CMGPD-SC to compare the death rates of settlers in Shuangcheng according to whether their families originated in urban Beijing and its environs or from rural northeast China. She found that descendants of migrants from Beijing experienced a persistent mortality disadvantage even though state policies privileged them. Campbell and Lee (2009) used the CMGPD-LN to study how household context affected mortality in adulthood and old age. They found that men who had lost their mothers in childhood or whose mothers were 35 or older when they were born had higher death rates in adulthood, and that men experienced elevated mortality risks in old age if they were born after a short preceding birth interval, to women who were 35 or older, to a father listed as disabled, or to a father who held a salaried official position. Dong and Lee (2014) used the CMGPD-LN to examine mortality in later life of men who had migrated from one village

to another in childhood and found that they had more favourable outcomes if they had kin in their destination village. Most recently, Zang and Campbell (2018) used the CMGPD-LN to investigate how co-residence with grandparents in childhood influenced mortality in adulthood and old age.

4.1.3 MARRIAGE AND HOUSEHOLD

Marriage has been a fruitful area for study because marriage timing and the overall chances of marriage closely reflected household priorities and individual privilege within the household. Marriage was the direct result of an explicit decision by the household about when a son or daughter would marry, and who they would marry. By contrast, with the exception of infanticide, fertility and mortality were outcomes that were influenced by household priorities and decisions but were also subject to a variety of other unrelated influences, making associations much more difficult to interpret. High status males were more likely to marry and if widowed, remarry. The first demonstration of the positive association between male socioeconomic and household status and marriage chances was for Daoyi in Lee and Campbell (1997, pp. 133–156, pp. 177–195). In the Imperial Lineage, social status was also positively associated with male marriage chances (Lee, Wang, & Ruan, 2001). Socioeconomic status of distant but co-residing kin influenced male marriage chances, and there was also clear evidence of sequencing among unmarried males of the same generation within the household (Campbell & Lee, 2008b). Higher status females tended to marry later but almost all females married eventually (Chen, Campbell, & Lee, 2014). Remarriage chances were tied to socioeconomic status as well, with higher status widowers more likely to remarry.

We also considered other aspects of marriage, including polygyny and the effects of economic shocks. Even though polygyny was one of the most widely noted features of marriage in China before the 20th century, it was extremely rare in the rural populations covered by the CMGPD-LN and CMGPD-SC. Even in the elite Imperial Lineage, polygyny became steadily less common over time, so that by the last half of the 19th century it was rare except among close relatives of the emperor. Moreover, polygyny was used primarily to extend the reproductive span of males rather than to father children with different partners at the same time (Lee et al., 2001). In rural Liaoning, economic hardship as reflected in elevated grain prices did not have an immediate effect on marriage the way it did on mortality and fertility, but had a lagged effect because elevated female infant and children mortality disproportionately reduced the numbers of girls reaching adulthood two decades later, worsening the imbalance in the marriage market (Campbell & Lee, 2008a).

Recently we have examined assortative marriage, that is who marries whom, for insight into family preferences regarding their affinal connections. This helps delineate social, economic and institutional boundaries between groups in historical China. Our first paper on the topic examined interethnic marriage in the CMGPD-SC for insight into whether in a unique institutional setting where Han and Manchu were allowed to intermarry without being affected by rules forbidding marriage between affiliates of the Eight Banners and regular civilians, they would do so (Chen, Campbell, & Dong, 2018). We found that marriage between Manchu and Han was common and that its likelihood depended on family characteristics including a family history of intermarriage, local marriage market composition, and other factors. Our second paper on the topic examined assortative mating by education and family class label in rural Shanxi in China in the middle of the 20th century (Xing et al., 2020) and found that both were important in marriage formation, and that patterns changed little before and after 1949, when the People's Republic of China was established. This was a novel finding because while there are many studies of educational assortative marriage in China in the last half of the 20th century, there are fewer that consider the middle of the 20th century and simultaneously consider the role of class labels.

Another line of work investigates household dynamics, including the growth of households and the formation of new ones by household division. In Liaoning, a large share of the population lived in large households with many distantly related individuals living together (Lee & Campbell, 1997, pp. 105–132; Lee & Gjerde, 1986). These households were highly hierarchical, with status and privilege determined by relationship to the household head (Lee & Campbell, 1997, pp. 133–156). The head and his or her children and grandchildren of the head were most privileged, and more distant kin were less privileged. When households divided, it was typically upon the death of a head or other senior relative whose presence linked different kin groups within the household (Lee & Campbell, 1998). Household heads were predominantly male, but widows sometimes inherited the headship after the death of their husband. Household division was a liberating experience for distant relatives of the

head who previously had been at the bottom of the hierarchy but as a result of division now exercised control over the resources of the newly formed household (Campbell & Lee, 1999).

4.2 COMPARISONS

For twenty years from 1994 to 2014, together with Feng Wang, we analysed the CMGPD datasets as part of a collaborative international comparison: the Eurasia Project in Population and Family History. By carrying out nearly identical analyses with similar data from different settings in Europe and Asia, we compared patterns of demographic responses to economic conditions in southwestern Sweden, northeast Japan, northeast China, eastern Belgium, and northern Italy. We interacted extensively with Tommy Bengtsson, Christer Lundh and Martin Dribe who worked on Scania in southwest Sweden, Akira Hayami, Noriko Tsuya, and Satomi Kurosu who worked on Fukushima in northeast Japan, Michel Oris and George Alter who worked on eastern Belgium, and Marco Breschi, Matteo Manfredini, and Renzo Derosas who worked on northern Italy.³⁵

The resulting comparisons of mortality (Bengtsson et al., 2004), fertility (Tsuya et al., 2010) and nuptiality (Lundh et al., 2014) revealed unanticipated similarities between East and West, the role of household context in shaping demographic behaviour, as well as unexpected differences. Key findings were that in the West, socioeconomic differences were important for shaping demographic responses to economic shocks, while in the East, socio-political differences in household context were more important. Overall, and in contrast with expectations based on Malthusian interpretations of population dynamics, demographic responses to economic shocks were weaker in the East than in the West. The emphasis on comparison by analysis of results from models that were the same across all the different datasets distinguished this effort from previous international comparisons of population and family in past times and led to novel results.

For insight into the strengths and weaknesses of the registers that were the basis of the CMGPD-LN, we also conducted comparisons of the same families recorded in the CMGPD-LN and their own genealogies. Among the materials we collected from each village we visited during our eight field trips were lineage genealogies. We transcribed these into a dataset and then compared the recording of lineage members between the CMGPD-LN and the genealogies (Campbell & Lee, 2002a). We found that as is already widely known, sons who died in infancy and childhood as well as daughters tended to be omitted from family genealogies, leading fertility estimated from genealogies to be underestimated. We also showed that fertility estimated from genealogies could be underestimated because they were more likely to omit adults who never married and married adults who did not have any surviving heirs. Whereas previous research had assumed that because of the omission of sons who died early and most daughters, fertility estimates from genealogies could be 'corrected' with an adjustment for infant and child mortality and the sex ratio at birth, the countervailing biases associated with the omission of childless adults made adjustment much more difficult, or perhaps even impossible.

We have initiated a new comparative, collaborative study of family and demographic behaviour in historical East Asia led by Hao Dong. Hao Dong harmonized datasets from northeast China, northeast Japan, Korea and Taiwan, and worked with Satomi Kurosu (Japan), and Wenshan Yang (Taiwan) on the analysis of these data. For these comparisons we have also made use of triennial Korean household registers from the county of Tansong made publicly available by a group of historians who at the time were mostly at Sungkyunkwan University and which we have turned into a longitudinal dataset through nominative linkage.³⁶ The resulting comparative studies explore how family context including presence and absence of various kin influenced demographic outcomes across East Asian populations (Dong et al., 2015a, 2015b; Dong, 2016; Dong et al., 2017).

4.3 SOCIAL MOBILITY, INEQUALITY AND MIGRATION

Our study of social mobility progressed from the analysis of associations in the outcomes of fathers and sons to the study of the role of networks of kin in shaping individual outcomes and finally to the study of lineages in their own right, with lineage membership as a key stratifying variable in historical

35 Each of the teams had other participants with whom we interacted more sporadically.

36 Cameron Campbell and Hao Dong wrote software to produce longitudinal links of records of the same individual in different registers to transform the cross-sectional data into a panel dataset (Dong et al., 2015b). The longitudinal links are available at <https://doi.org/10.14711/dataset/IVIDZV>.

Chinese society. Initial studies of father-son associations revealed that a son had a greater chance of obtaining a salaried official position in Daoyi if his father held one (Lee & Campbell, 1997, pp. 196–215). Comparison with results of studies of social mobility in 19th century North America and Europe revealed that the attainment advantages of the sons of locally elite fathers were nevertheless much less pronounced in Liaoning than in the West (Campbell & Lee, 2003). Ethnic mobility accompanied social mobility in the sense that Han men who held salaried official positions were more likely to change from Han to Manchu names (Campbell, Lee, & Elliott, 2002). In every generation, large proportions of the men in Liaoning who entered the local elite by obtaining salaried government positions not only were 'new' in the sense that not only did their father not hold a position, but neither did any of their other patrilineal relatives (Campbell & Lee, 2003). Positions held by other kin were usually a source of advantage, though not always (Campbell & Lee, 2008b).

Lineage membership was also a source of differentiation in rural Liaoning. Social and demographic outcomes, especially attainment chances and marriage chances, depended not only on individual and household characteristics but also lineage affiliation (Campbell & Lee, 2008c). There was long-term continuity not only during the Qing but between the Qing and the late 20th century in the relative status of lineages (Campbell & Lee, 2011). Socioeconomic privilege not only increased the number of children a man had, but increased the total number of descendants he had for as many as six generations, meaning that in every generation, a disproportionate share of the population was descended from the most socioeconomically privileged members of the population in previous generations (Song, Campbell, & Lee, 2015). We have also explored computational approaches to the study of lineages: Fu et al. (2018) used visualization and network techniques to study the determinants of the morphology of descent line structure.

Newer work considers inequality from an even broader perspective. Chen (2017) examines stratification based on institutional affiliation and landholding in Shuangcheng. The state specified different entitlements to land according to population category defined by institutional affiliation. These differential land entitlements affected landholding as well as access to other social and economic privileges. The residents of Shuangcheng challenged the state defined social hierarchy in some cases but at the same time reinforced it in others. Noellert (2020) examines individual-level data on Land Reform events in Shuangcheng after 1945 and found that a redistribution of power away from local strongmen paved the way for the reallocation of property, which continued to be defined by state entitlements.

We have also conducted studies of migration. The CMGPD-LN follows households when they move within Liaoning. When people leave the region entirely, usually illegally, that is recorded as well. Our first study examined the determinants of legal migration of households within Liaoning and illegal departure from the region (Campbell & Lee, 2001). Household age structure conditioned legal migration: 'younger' households with fewer elderly dependents were much more likely to migrate. Households with men who held salaried positions, meanwhile, were less likely to move. Illegal departure was more common for men who were unmarried or widowed, distant relatives of the household head, or members of smaller households. Dong et al. (2015a) compares patterns in northeast China with those in 18th and 19th century Korea and Japan.

4.4 SOCIAL AND SPATIAL ORIGINS OF UNIVERSITY STUDENTS IN 20TH CENTURY CHINA

Studies based on CUSD datasets of student registration cards and other materials have illuminated shifts in the spatial and social origins of university students in China from the late 19th century to the beginning of the 21st century. Whereas during the Qing educated elites were recruited nationally via the examination system until its abolition in 1905, the educated elites who dominated Republican China in the first half of the 20th century were generally drawn from merchant and white-collar professional families in the major coastal cities (Liang et al., 2017; Ren et al., 2020). Liang et al. (2013) moreover showed that immediately after 1949, student origins at Peking University and Soochow University continued to resemble Republican-era universities, with disproportionate numbers of students coming from business and professional families in the coastal cities.

More importantly, Liang et al. (2012, 2013) also show that the introduction of standardized exams (*gaokao*) in 1955 together with a major expansion in primary and secondary education fundamentally transformed the composition of university-eligible students. In particular, the numbers of students from

farm and factory families who were the first in their families to attend college significantly increased. This pattern persisted well into the 1990s, when the share of children of professionals began to rise again. At least until 2004, however, approximately 30% of students in Peking University and 40% of students in Soochow University still originated from working-class families. This was a very different pattern from much of the West, where the students who attend the elite private universities that are the counterpart of Peking and Soochow Universities overwhelmingly come from high-income families. These findings had considerable impact on ongoing Chinese debates at the turn of the twenty-first century whether the college entrance examinations still maintained earlier opportunities for students from families of modest means, or favoured students from already well-off families.³⁷

4.5 QING OFFICIALDOM AND THE CAREERS OF OFFICIALS

Analysis of the CGED-Q has already yielded insights into Qing officialdom and the careers of officials not available from traditional approaches to the study of the Qing civil service which emphasize case studies of individuals or offices, or specific time periods. Ren et al. (2016), Chen (2019) and Chen et al. (2020) show that the central government, especially its upper reaches, were dominated by Manchu and other bannermen right up to the end of the Qing. Only a relatively small share of Han who qualified by their civil service examination performance served in the central government largely confined to the Hanlin Academy and related offices. Outside the central government, however, officials were predominantly Han, and included more holders of purchased degrees than holders of exam degrees. Median career length was just under seven years, except for bannermen and holders of *gongsheng* exam degrees, whose medium career length was three years (Chen et al., 2020). The abolition of the examination system in 1905 had little effect on the holders of exam degrees who were already officials, or on holders of exam degrees who awaited appointment. Chen, Campbell and Lee (2018) examined Banner officials at the very end of the Qing and found that their numbers and positions changed little during the New Government period, but that their share of officials declined because of an increase in the number of Han officials. Campbell (2020) shows that after the abolition of the civil service exams in 1905, men who already held exam degrees continued to be appointed at the same pace as before, and the turnover of exam degree holders who were officials was unaffected. Such results challenge claims made in other studies that the abolition of the examinations adversely affected aspiring elites (Bai & Jia, 2016).

5 CONCLUSION

Looking back at four decades of collaboration on the study of demographic, social and economic history, we have some reflections and observations. The first is that we have been extraordinarily fortunate in terms of finding, acquiring, and constructing the diverse, large microdata sources which are the basis of almost all our research. This has been a group effort. Our achievements in identifying new sources of microdata to understand China's past and sometimes present are increasingly due to collaborations with our colleagues in the Lee-Campbell Research Group. Notable examples are Chen Liang and Bamboo Ren's compilations of Republican-era student registration records, Yuxue Ren sharing with Cameron Campbell her ongoing work on the *jinshenlu*, Matthew Noellert and Xiangning Li's discovery of new materials on China's rural reconstruction in the mid 20th century, and recent discoveries by Bamboo Y. Ren, Yibei Wu, and Li Yang, of materials on overseas students and educated professionals.

Second, institutional support was crucial for us to acquire the microdata we discuss in this article. Institutions that very generously made materials from their holdings available include the Liaoning Provincial Archives, the First Historical Archives, the Liaoning Local History Office, the Genealogical

37 An initial article publicizing these results in 2012 (Liang et al., 2012) inspired immediate and widespread public debate, summarized on our group web pages, with over 100 response articles, reposted, reprinted or rebroadcast on over 1000 Chinese media sites. These included several televised reports and discussions, most notably a widely publicized recommendation of our subsequent book, Liang et al. (2013), made by CPPCC Chairman Zhengsheng Yu during a 2014 meeting on education held in conjunction with the ongoing Chinese People's Political Consultative Congress. More recently, an extended video commentary on the book by Yu Liang posted on the Chinese website Bilibili on July 12, 2020 had been viewed 240,000 times and liked 22,000 times by July 15, 2020.

Society of Utah (now called [FamilySearch](#)), the Shanxi University Research Centre for Chinese Social History, the Shuangcheng County Archives and numerous universities in China and the United States, such as Peking University, Soochow University, Shanghai Jiao Tong University, Tsinghua University, Zhejiang University, and Columbia University. For other datasets, most notably the CGED-Q, we relied on materials that had been published or made publicly available for download. The most important of these is the Tsinghua University Library collection of *jinshenlu* published as *Qingdai Jinshenlu Jicheng* that along with *jinshenlu* editions made available by the Harvard Yenching Library and Columbia University Library were the basis of most of the CGED-Q.

Third, we also benefitted from many other individuals who contributed to our construction of these datasets. Listing everyone here would be impossible, but we can single out some individuals who played especially important roles. Ju Deyuan, Robert Eng, Alice Suen, Anna Chi and others helped James Lee initiate research on Daoyi. Mel Thatcher arranged for access to the collections of the Genealogical Society of Utah. Ts'ui-jung Liu, Sufen Liu, and Huimin Lai at the Academia Sinica facilitated the creation of the CMGPD-IL and one of the CMGPD-LN register series. Among the coders at the Academia Sinica, Shu-mei Tsay made the most contributions to the CMGPD-IL and CMGPD-LN. Shuang Chen, Matt Noellert, and Bijia Chen coordinated and oversaw the data entry of the CMGPD-SC, CRRD-LR and CRRD-SQ, and CGED-Q, respectively. Similarly, Liang Chen, Bamboo Ren, Hao Zhang, Yibei Wu, and Li Yang initiated, coordinated, and oversaw the creation of various subsets of the CUSD and CPOD. Hao Dong helped create longitudinal links for the Korean registers and led the efforts to harmonize the CMGPD and other datasets from Japan, Korea, and Taiwan. Many, many coders worked tirelessly to enter all these data. It is not possible to list all of them here, but we highlight six who made especially large contributions over extended periods of time: Huicheng Sun, Jiyang, and Xing Xiao entered much of the CMGPD-LN, CMGPD-SC, CRRD-LR, and together with Xiaodong Ge, Yibei Liu, and Mi Zhao, the CGED-Q.

Fourth, we could not have proceeded without generous institutional and occasionally personal support. James Lee began his career at the California Institute of Technology and Cameron Campbell met him there while an undergraduate. In retrospect, Caltech was one of only a few places that in the early 1980s would support an assistant and then associate and finally full professor of humanities/history to carry out quantitative research on China. It is also hard to imagine that at any other institution, a sophomore studying electrical engineering who had a side interest in Chinese history but no language ability could walk in to a history professor's office and after some discussions outline a plan to reorganize data management and analysis for an ongoing project and then become a collaborator. In graduate school at Penn and then as an assistant, associate and then full professor and sociology at UCLA, Campbell was supported by mentors and then colleagues even though his work was esoteric.

Internal funding and administrative support from the California Institute of Technology, the University of Michigan, UCLA, Peking University, the Hong Kong University of Science and Technology, Shanghai Jiao Tong University, and from other universities in China allowed us to greatly expand our data acquisition and dataset construction, as well as to apply for sustained research support from the National Institutes of Health in the USA, the National Science Council in Taiwan, the National Natural Science Foundation in mainland China, and the Research Grants Council in Hong Kong. Equally important, these universities provided opportunities and sometimes funding to collaborate with the graduate students, postdoctoral fellows, and visiting professors who made crucial contributions to, and in some cases led, the data construction and research which define much of our work over the last forty years and hopefully for many years in the future. We are especially grateful to Myron Guttman, who as head of the Inter-university Consortium for Political and Social Research, provided guidance and support when we were scaling up our operations and seeking for the first time substantial extramural funding.

Long-term collaborations played a key role in advancing our work. The most sustained and for us influential collaboration was the twenty years we spent working with colleagues from a variety of countries and disciplines on the Eurasia Project in Population and Family History. Interactions with project participants stimulated us to broaden our range of research topics, learn and apply more advanced methods, and seek opportunities for comparisons for our other projects. The camaraderie that grew out of frequent, sustained interaction with others working with data like ours and on similar topics was also important for our own morale. We have especially fond memories of our fruitful two decade-long collaboration with Feng Wang which produced a number of discrete studies as well as Lee and Wang (1999) and Tsuya et al. (2010) and our collaborations with Tommy Bengtsson and Noriko Tsuya, which involved reciprocal visits by us in Lund and Tokyo and by them in Pasadena.

Many short-term collaborations on specific papers and long-term projects were also important. Yizhuang Ding and Songyi Guo advised on the CMGPD-LN and with the help of Gao Jing from the Liaoning Provincial Gazetteer Office we conducted fieldwork with them that resulted in Ding et al. (2004). Songyi Guo also shared his expertise with us when we were constructing and analysing the CMGPD-IL. We have been fortunate to co-author with a variety of others on papers or sets of papers using our datasets, including Lawrence Anthony, Mark Elliott, Robert Eng, William Lavelly, Chris Myers, Xi Song, Alice Suen, Guofu Tan, Emma Zang, and Siwei Fu and other members of Huamin Qu's group. Similarly, we have benefited from interaction with Akira Hayami, Kuentae Kim, Satomi Kurosu, Sangkuk Lee, Ts'ui-jung Liu, Byun-giu Son, Noriko Tsuya, Wenshan Yang, and other collaborators.

Looking back, we believe that a distinctive feature of our research that has been central to our success has been our inductive, data-driven approach which emphasizes discovery of facts about demographic behaviour and family, social and economic organization through empirical analysis of the datasets we have constructed. We have always started with data that we thought might help us investigate a topic of general interest, and then through exploratory and descriptive analysis sought to uncover key patterns of demographic behaviour and family and social organization, moving to carefully specified regression-based models only after extensive work to verify the data and then elaborate on relationships and patterns discovered in descriptive analysis. While this approach is time-consuming, sometimes taking years to locate, access, enter and clean data for the analyses that led to our major results, we believe that the result has been a fundamental transformation in our understanding of basic patterns of family, social and economic organization in China in the past and increasingly up to the very recent present.

Our strong belief in the importance of a scholarship of microdata-driven empirical discovery as opposed to a scholarship of interpretation is because there is much about the Chinese past we do not know, or worse, think we know, but are wrong.³⁸ Whenever feasible and permissible, therefore, we have coded datasets in their entirety and devoted considerable time and energy to produce detailed documentation and User Guides to accompany possible future public data releases, and create a complete, permanent resource to be used by ourselves and others to study a wide range of topics. Related projects described in this special issue are underway for a variety of other locations around the world and we look forward to a future where such datasets are produced and used routinely in social science and in history to discover basic facts about life in the past, right up to the time a few decades ago when longitudinal surveys and other sources become available.

ACKNOWLEDGEMENTS

While working on this manuscript, Lee and Campbell received support from HK RGC GRF 16602117 (Lee PI), HK RGC GRF 16601718 (Campbell PI), and HK RGC GRF 16600017 (Campbell PI). We are grateful to the members of the Lee-Campbell Group for their comments, corrections, and other feedback.

REFERENCES

- Bai, Y., & Jia, R. (2016). Elite recruitment and political stability: The impact of the abolition of China's civil service exam system. *Econometrica*, 84(2), 677–733. doi: [10.3982/ECTA13448](https://doi.org/10.3982/ECTA13448)
- Bengtsson, T., Campbell, C. D., Lee, J. Z. et al. (2004). *Life under pressure: Mortality and living standards in Europe and Asia, 1700–1900*. Cambridge, MA: MIT Press.

38 A good example world-wide would be Malthus' previously commonly accepted description of the Chinese demographic system and its consequences (Lee & Wang, 1999). Another largely Chinese example would be the debate over opportunity for first-college students from families of marginal means to attend some of China's most competitive universities. See too our acknowledgement in Bengtsson et al. (2004) of our own past misunderstandings of mortality and related behavior and our description of the processes by which we were able to correct ourselves and achieve new understandings of the past (435–439).

- Cai, S., Lee, J. Campbell, C. D. & Myers, C. (1994). Zongrenfu dang'an diannaoku de jianli, fenxi, liyong, jiqi kunnan [Machine analysis and data coding of the Qing imperial lineage]. In J. Lee & S. Guo (Eds.), *Qingdai huangzu renkou xingwei he shehui huanjing [The Qing imperial lineage: Social structure and population behavior]* (pp. 191–204). Beijing: Peking University Press.
- Campbell, C. D. (1997). Public health efforts in China before 1949 and their effects on mortality: The case of Beijing. *Social Science History*, 21(2), 179–218. doi: [10.2307/1171273](https://doi.org/10.2307/1171273)
- Campbell, C. D. (2001). Mortality change and the epidemiological transition in Beijing, 1644–1990. In T.-j. Liu, J. Lee, D. S. Reher, O. Saito, & F. Wang (Eds.), *Asian population history* (pp. 221–247). Oxford: Oxford University Press.
- Campbell, C. D. (2016). Large-scale collaboration and comparison in historical demography: Reflections on the Eurasia project. In K. Matthijs, S. Hin, J. Kok, & H. Matsuo (Eds.), *The future of historical demography: Upside and inside out* (pp. 193–196). Leuven: Acco.
- Campbell, C. D. (2020). Qingmo keju tingfei dui shiren wenguan qunti de yingxiang-jiyu weiguan dashuju de hongguan xin shijiao [The influence of the abolition of the examinations at the end of the Qing on the holders of exam degrees]. *Shehui kexue jikan [Social Science Journal]*, 4(249), 156–166.
- Campbell, C. D., Dong, H., & Lee, J. Z. (2016). *China multigenerational panel datasets training guide*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.
- Campbell, C. D., Wang, F., & Lee, J. Z. (2002). Pretransitional fertility in China. *Population and Development Review*, 28(4), 735–750. doi: [10.1111/j.1728-4457.2002.00735.x](https://doi.org/10.1111/j.1728-4457.2002.00735.x)
- Campbell, C. D., & Lee, J. Z. (1996). A death in the family: Household structure and mortality in rural Liaoning: Life-event and time-series analysis, 1792–1867. *The History of the Family*, 1(3), 297–328. doi: [10.1016/S1081-602X\(96\)90026-3](https://doi.org/10.1016/S1081-602X(96)90026-3)
- Campbell, C. D., & Lee, J. Z. (1999). Causes and consequences of household division in northeast China, 1789–1909. In Y. Ding, S. Guo, & J. Lee (Eds.), *Hunyin yu jiating: Dongxi bijiao shi [Marriage and family, East-West comparative methods]* (pp. 1–32). Beijing: Peking University Press.
- Campbell, C. D., & Lee, J. Z. (2000). Price fluctuations, family structure and mortality in two rural Chinese populations: Household responses to economic stress in eighteenth- and nineteenth-century Liaoning. In T. Bengtsson & O. Saito (Eds.), *Population and the economy: From hunger to modern economic growth* (pp. 371–420). Oxford: Oxford University Press.
- Campbell, C. D., & Lee, J. Z. (2001). Free and unfree labour in Qing China: Emigration and escape among the bannermen of northeast China, 1789–1909. *The History of the Family*, 6(4), 455–476. doi: [10.1016/S1081-602X\(01\)00088-4](https://doi.org/10.1016/S1081-602X(01)00088-4)
- Campbell, C. D., & Lee, J. Z. (2002a, publ. 2006). State views and local views of population: Linking and comparing genealogies and household registers in Liaoning, 1749–1909. *History and Computing*, 14(1-2), 9–29.
- Campbell, C. D., & Lee, J. Z. (2002b). When husbands and parents die: Widowhood and orphanhood in late imperial Liaoning, 1789–1909. In R. Derosas & M. Oris (Eds.), *When dad dies* (pp. 313–334). Bern: Peter Lang.
- Campbell, C. D., & Lee, J. Z. (2003). Social mobility from a kinship perspective: Rural Liaoning, 1789–1909. *International Review of Social History*, 48(1), 1–26. doi: [10.1017/S0020859002000901](https://doi.org/10.1017/S0020859002000901)
- Campbell, C. D., & Lee, J. Z. (2004). Mortality and household in seven Liaodong populations. In T. Bengtsson, C. D. Campbell, J. Z. Lee et al., *Life under pressure: Mortality and living standards in Europe and Asia, 1700–1900* (293–324). Cambridge, MA: MIT Press.
- Campbell, C. D., & Lee, J. Z. (2008a). Economic conditions and male first marriage in northeast China, 1749–1909. *Sungkyun Journal of East Asian Studies*, 8(1), 17–42. doi: [10.21866/esjeas.2008.8.1.002](https://doi.org/10.21866/esjeas.2008.8.1.002)
- Campbell, C. D., & Lee, J. Z. (2008b). Kin networks, marriage, and social mobility in late imperial China. *Social Science History*, 32(2), 175–214. doi: doi.org/10.1017/S0145553200010749
- Campbell, C. D., & Lee, J. Z. (2008c). Villages, descent groups, households and individual outcomes in rural Liaoning, 1789–1909. In T. Bengtsson & G. Mineau (Eds.), *Kinship and demographic behaviour in the past* (pp. 73–104). Springer.
- Campbell, C. D., & Lee, J. Z. (2009). Long-term mortality consequences of childhood family context in Liaoning, China, 1749–1909. *Social Science and Medicine*, 68(9), 1641–1648. doi: [10.1016/j.socscimed.2009.02.017](https://doi.org/10.1016/j.socscimed.2009.02.017)
- Campbell, C. D., & Lee, J. Z. (2010a). Demographic impacts of climatic fluctuations in northeast China, 1749–1909. In S. Kurosu, T. Bengtsson, & C. D. Campbell (Eds.), *Demographic responses to economic and environmental crisis* (pp. 107–132). Kashiwa: Reitaku University Press.

- Campbell, C. D., & Lee, J. Z. (2010b). Fertility control in historical China revisited: New methods for an old debate. *History of the Family*, 15(4), 370–385. doi: [10.1016/j.hisfam.2010.09.003](https://doi.org/10.1016/j.hisfam.2010.09.003)
- Campbell, C. D., & Lee, J. Z. (2011). Kinship and the long-term persistence of inequality in Liaoning, China, 1749–2005. *Chinese Sociological Review*, 44(1), 71–104. doi: [10.2753/CSA2162-0555440105](https://doi.org/10.2753/CSA2162-0555440105)
- Campbell, C. D., Lee, J. Z., & Elliott, M. (2002). Identity construction and reconstruction: Naming and Manchu ethnicity in northeast China, 1749–1909. *Historical Methods*, 35(3), 101–116. doi: [10.1080/01615440209601201](https://doi.org/10.1080/01615440209601201)
- Chen, B. (2019). *Origins and career patterns of the Qing government officials (1850–1912): Evidence from the China Government Employee Dataset-Qing (CGED-Q)* (PhD dissertation). Hong Kong University of Science and Technology Division of Social Science, China
- Chen, B., Campbell, C. D., & Dong, H. (2018). Interethnic marriage in northeast China, 1866–1913. *Demographic Research*, 38(34), 929–966. doi: [10.4054/DemRes.2018.38.34](https://doi.org/10.4054/DemRes.2018.38.34)
- Chen, B., Campbell, C. D., & Lee, J. Z. (2018). Qingmo xingzheng qianhou qiren yu zongshi guanyuan de guanzhi bianhua chutan-yi jinshenlu shujuku wei cailiao de fenxi [The transition of banner and imperial lineage officials during the late Qing reform period: Evidence from the Qing Jinshenlu Database]. *Qingshi Yanjiu [Studies in Qing History]*, 2018(4), 10–20.
- Chen, B., Campbell, C. D., Ren, Y., & Lee, J. Z. (2020). Big data for the study of Qing officialdom: The China government employee database-Qing (CGED-Q). *The Journal of Chinese History*, 4(2):431–460. doi: [0.1017/jch.2020.15](https://doi.org/0.1017/jch.2020.15)
- Chen, S. (2009). *Where urban migrants meet rural settlers: State categories, social boundaries, and wealth stratification in northeast China, 1815–1913* (PhD Dissertation). University of Michigan, Department of History, USA.
- Chen, S. (2017). *State-sponsored inequality. The banner system and social stratification in northeast China*. Stanford: Stanford University Press.
- Chen, S., Campbell, C. D., & Lee, J. Z. (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.
- Chen, S., Campbell, C. D., & Lee, J. Z. (2014). Categorical inequality and gender difference: Marriage and remarriage in northeast China, 1749–1913. In C. Lundh, S. Kurosu et al. (Eds.), *Similarity in difference: Marriage in Europe and Asia, 1700–1900* (pp. 393–438). Cambridge, MA: MIT Press
- Chen, S., Lee, J. Z., & Campbell, C. D. (2010). Wealth stratification and reproduction in northeast China, 1866–1907. *History of the Family*, 15(4), 386–412. doi: [10.1016/j.hisfam.2010.10.001](https://doi.org/10.1016/j.hisfam.2010.10.001)
- Ding, Y., Guo, S., Lee, J. Z., & Campbell, C. D. (2004). *Liaodong yimin zhong de qiren shehui [Banner society and the settlement of eastern Liaodong]*. Shanghai: Shanghai shehui kexue chubanshe.
- Dong, H. (2016). *Patriarchy, family system and kin effects on individual demographic behaviour throughout the life course: East Asia, 1678–1945* (PhD dissertation). Hong Kong University of Science and Technology Division of Social Science, China.
- Dong, H., Campbell, C. D., Kurosu, S., & Lee, J. Z. (2015a). Household context and individual departure: The case of escape in three 'unfree' East Asian populations, 1700–1900. *Chinese Journal of Sociology*, 1(4), 513–539. doi: [10.1177/2057150X15614547](https://doi.org/10.1177/2057150X15614547)
- Dong, H., Campbell, C. D., Kurosu, S., Yang, W., & Lee, J. Z. (2015b). New sources for comparative social science: Historical population panel data from East Asia. *Demography*, 52(3), 1061–1088. doi: [10.1007/s13524-015-0397-y](https://doi.org/10.1007/s13524-015-0397-y)
- Dong, H., & Lee, J. Z. (2014). Kinship matters: Long-term mortality consequences of childhood migration, historical evidence from northeast China, 1792–1909. *Social Science and Medicine*, 119, 274–283. doi: [10.1016/j.socscimed.2013.10.024](https://doi.org/10.1016/j.socscimed.2013.10.024)
- Dong, H., Manfredini, M., Kurosu, S., Yang, W., & Lee, J. Z. (2017). Kin and birth order effects on male child mortality: Three East Asian populations, 1716–1945. *Evolution and Human Behaviour*, 38(2), 208–216. doi: [10.1016/j.evolhumbehav.2016.10.001](https://doi.org/10.1016/j.evolhumbehav.2016.10.001)
- Fu, S., Dong, H., Cui, W., Zhao, J., & Qu, H. (2018). How do ancestral traits shape family trees over generations? *IEEE Transactions on Visualization and Computer Graphics*, 24(1), 205–214. doi: [10.1109/TVCG.2017.2744080](https://doi.org/10.1109/TVCG.2017.2744080)
- Gautier, E., & Henry, L. (1958). *La population de Crulai, paroisse normande: Étude historique*. Travaux et documents – Institut national d'études démographiques, vol. 33. Paris: Presses universitaires de France.
- Ho, P.-t. (1959). *Studies on the Population of China, 1368–1953*. Harvard East Asian Series 4. Cambridge, MA: Harvard University Press.

- Ju, D. (1994). Qingdai huangzu renkou ciji [Demographic records of the Qing imperial lineage]. In J. Z. Lee & S. Guo (Eds.), *Qingdai huangzu renkou xingwei he shehui huanjing [The Qing imperial lineage: Social structure and population behaviour]* (pp. 170–190). Beijing: Peking University Press.
- Laslett, P., & Wall, R. (1972). *Household and family in past time: Comparative studies in the size and structure of the domestic group over the last three centuries in England, France, Serbia, Japan, and colonial North America, with further materials from western Europe*. New York: Cambridge University Press.
- Lavelly, W., Lee, J. Z., & Wang, F. (1990). Chinese demography: The state of the field. *Journal of Asian Studies*, 50(1), 807–834. doi: <https://doi.org/10.2307/2058237>
- Lee, J. Z., Anthony, L., & Suen, A. (1988). Liaoning sheng chengren siwang lu, 1796–1819 [Adult mortality in Liaoning, 1796–1819]. In *Qingzhu diyi lishi dang'an guan liushi zhounian lunwen ji [Proceedings of the Symposium on the Occasion of the Sixtieth Anniversary of the First Historical Archives]*, Volume 2 (pp. 885–898). Beijing: Zhonghua Shuju.
- Lee, J. Z., & Eng, R. (1984). Population and family history in eighteenth-century Manchuria: Preliminary results from Daoyi 1774–1798. *Ch'ing-shih wen-t'i*, 5(1), 1–55.
- Lee, J. Z., & Campbell, C. D. (1997). *Fate and fortune in rural china: Social organization and population behaviour in rural Liaoning, 1774–1873*. Cambridge: Cambridge University Press.
- Lee, J. Z., & Campbell, C. D. (1998). Headship succession and household division in three Chinese banner serf populations, 1789–1909. *Continuity and Change*, 13(1), 117–141. doi: [10.1017/S0268416098003063](https://doi.org/10.1017/S0268416098003063)
- Lee, J. Z., Campbell, C. D., & Anthony, L. (1995). A century of mortality in rural Liaoning, 1774–1873. In S. Harrell (Ed.), *Chinese Historical Microdemography* (pp. 163–182). Berkeley: University of California Press.
- 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.
- Lee, J. Z., Campbell, C. D., & Wang, F. (1993). An introduction to the demography of the Qing imperial lineage, 1644–1911. In R. Schofield & D. Reher (Eds.), *Old and new methods in historical demography* (pp. 361–382). Oxford: Oxford University Press.
- Lee, J. Z., Campbell, C. D., & Wang, F. (2002). Positive checks or Chinese checks? *Journal of Asian Studies*, 61(2), 591–607. doi: [10.2307/2700301](https://doi.org/10.2307/2700301)
- Lee, J. Z., Campbell, C. D., & Tan, G. (1992). Infanticide and family planning in rural Liaoning, 1774–1873. In L. Li & T. Rawski (Eds.), *Chinese History in Economic Perspective* (pp. 149–176). Berkeley: University of California Press.
- Lee, J. Z., & Wang, F. (1999). *One quarter of humanity: Malthusian mythology and Chinese realities, 1700–2000*. Cambridge, Mass.: Harvard University Press.
- Lee, J. Z., Wang, F., & Campbell, C. D. (1994). Infant and child mortality among the Qing nobility: Implications for two types of positive check. *Population Studies*, 48(3), 395–412. doi: [10.1080/0032472031000147936](https://doi.org/10.1080/0032472031000147936)
- Lee, J. Z., Wang, F., & Ruan, D. (2001). Nuptiality among the Qing nobility: 1600–1900. In T.-j. Liu, J. Z. Lee, D. Reher, O. Saito, & F. Wang (Eds.), *Asian Historical Demography* (pp. 353–373). Oxford: International Studies in Demography, Oxford University Press.
- Lee, J. Z., & Gjerde, J. (1986). Comparative household morphology of stem, joint, and nuclear household systems: Norway, China, and the United States. *Continuity and Change*, 1(1), 89–111. doi: [10.1017/S0268416000000084](https://doi.org/10.1017/S0268416000000084)
- Lee, J. Z., & Guo, S. (Eds.). (1994). *Qingdai huangzu renkou xingwei he shehui huanjing [The Qing imperial lineage: Social structure and population behaviour]*. Beijing: Peking University Press.
- Lee, R., & Steckel, R. H. (2006). Life under pressure: An appreciation and appraisal. *Historical Methods*, 39(4), 171–176. Retrieved from https://u.demog.berkeley.edu/~rlee/oldceda/Publications/pdfs/rlee/LeeSteckel_LifeUnderPressureReview_HistoricalMethods2006.pdf
- Liang, C., Dong, H., Ren, Y., & Lee, J. Z. (2017). Jiangshan dai you rencai chu, geling fengsao shushinian [Changes in the social and geographical origins of China's educated elites, 1865–2014], *Shehuixue yanjiu [Sociology Research]*, 2017(3), 48–70.
- Liang, C., Lee, J. Z., Zhang, H., Li, L., Ruan, D., Campbell, C. D., & Yang, S. (2012). Wusheng de geming: Beijing daxue yu suzhou daxue xuesheng shehui laiyan yanjiu [Silent revolution: The social origins of Peking University and Soochow University undergraduates, 1952–2002]. *Zhongguo shehui kexue [Chinese Social Science]*, 2012(1), 99–119.

- Liang, C., Zhang, H., Li, L., Ruan, D., Campbell, C. D., & Lee, J. Z. (2013). *Wusheng de geming: Beijing daxue, Suzhou daxue xuesheng shehui lai yuan yanjiu, 1949–2002* [Silent revolution: Social origins of Peking university and Suzhou university students, 1949–2002]. Beijing: Sanlian Books.
- Liu, T.-j., Lee, J. Z., Reher, D., Saito, O., & Wang, F. (Eds.) (2001). *Asian Population History*. Oxford: International Studies in Demography, Oxford University Press.
- Lundh, C., Kurosu, S. et al. (2014). *Similarity in difference: Marriage in Europe and Asia, 1700–1900*. Cambridge, MA: MIT Press.
- Noellert, M. (2020). *Power over property: The political economy of communist land reform in China*. Ann Arbor: University of Michigan Press.
- Ren, B., Liang, C., & Lee, J. Z. (2020). Mutable inequalities: Meritocracy and the making of the Chinese Academe 1912–1952, A data analytic approach. *China Quarterly*. Accepted for publication.
- Ren, Y., Chen, B., Hao, X., Campbell, C. D., & Lee, J. Z. (2016). Qingdai jinshenlu lianghua shujuku yu guangliao qunti yanjiu [The Qing jinshenlu database: A new source for the study of Qing officials]. *Qingshi Yanjiu* [Qing History Research], 2016(4), 61–77.
- Ren, Y., Chen, B., Hao, X., Campbell, C. D., & Lee, J. Z. (2019). *Zhongguo lishi guanyuan lianghua shujuku-qingdai jinshenlu 1900–1912 shidian gongkaiban yonghu zhinan* [The China government employee database-Qing Jinshenlu 1900–1912 Public Release User Guide]. Hong Kong: The Hong Kong University of Science and Technology.
- Song, X., Campbell, C. D., & Lee, J. Z. (2015). Ancestry matters: Patrilineage growth and extinction. *American Sociological Review*, 80(3), 574–602. doi: [10.1177/0003122415576516](https://doi.org/10.1177/0003122415576516)
- Tsuya, N.O., Wang, F., Alter, G., Lee, J. Z. et al. (2010). *Prudence and pressure: Reproduction and human agency in Europe and Asia, 1700–1900*. Cambridge, MA: MIT Press.
- Wang, F., Campbell, C. D., & Lee, J. Z. (2010). Agency, hierarchies, and reproduction in northeastern China, 1789 to 1840. In N. Tsuya, F. Wang, G. Alter, J. Z. Lee et al., *Prudence and pressure: Reproduction and human agency in Europe and Asia, 1700–1900* (pp. 287–316). MIT Press.
- Wang, F., & Lee, J. Z. (1998). Adoption among the Qing nobility and its implications for Chinese demographic behaviour. *History of the Family*, 3(3), 411–428. doi: [10.1016/S1081-602X\(99\)80255-3](https://doi.org/10.1016/S1081-602X(99)80255-3)
- Wang, F., Lee, J. Z., & Campbell, C. D. (1995). Marital fertility control among the Qing nobility: Implications for two types of preventive check. *Population Studies*, 49(3), 383–400. doi: [10.1080/0032472031000148736](https://doi.org/10.1080/0032472031000148736)
- Wang, H., Chen, S., Dong, H., Noellert, M., Campbell, C. D., & Lee, J. Z. (2013). *China multi-generational panel dataset, Shuangcheng (CMGPD-SC) 1866–1914. User Guide*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.
- Wang, L. (2012). *Qingdai huangzu chengyuan de xianfu shenfen yu guanzhi huode, 1644–1911* [Ascribed status and official attainment by members of the Qing Imperial Lineage, 1644–1911] (PhD Dissertation). Peking University Department of Sociology, China
- Wang, L., Lee, J. Z., & Campbell, C. D. (2010). Institutions and inequality: Comparing the Zongshi and the Jueluo in the Qing imperial lineage. *Sungkyun Journal of East Asian Studies*, 10(1), 33–61. doi: [10.21866/esjeas.2010.10.1.003](https://doi.org/10.21866/esjeas.2010.10.1.003)
- Wolf, A., & Huang, C.-s. (1980). *Marriage and adoption in China, 1895–1945*. Stanford: Stanford University Press.
- Xing, L., Campbell, C. D., Li, X., Noellert, M., & Lee, J. Z. (2020). Education, class and assortative marriage in rural Shanxi, China in the mid-twentieth century. *Research in Social Stratification and Mobility*, 66(2020), 100460, 1–15. doi: [10.1016/j.rssm.2019.100460](https://doi.org/10.1016/j.rssm.2019.100460)
- Zang, E., & Campbell, C. D. (2018). Males' later life mortality consequences of coresidence with paternal grandparents: Evidence from northeast China, 1789–1909. *Demography*, 55(2), 435–457. doi: [10.1007/s13524-018-0653-z](https://doi.org/10.1007/s13524-018-0653-z)