State Views and Local Views of Population: Linking and Comparing Genealogical and Household Register Data in Liaoning, 1749-1909

Cameron Campbell
UCLA
camcam@ucla.edu

James Lee
Caltech/University of Michigan
jql@umich.edu

Prepared for the presentation at the session "Asian and Asian-American Families in Historical Perspective" at the Population Association of America, April 1-3, 2004, Boston MA.

Previous versions were presented at the session "New directions in linking historical demographic data sources" at the Social Science History Association meetings, Baltimore, MD, November 13-16, 2003, and the IMAG workshop of the North Atlantic Population Project, Montreal, November 10-11, 2003. We are grateful to the audiences for their comments and suggestions.

State Views and Local Views of Population:
Linking and Comparing Genealogical and Household Register Data in Liaoning, 1749-1909

Cameron Campbell (UCLA)

James Z. Lee (University of Michigan)

Introduction

Most micro-level quantitative studies of Chinese society rely on individual level data from locally produced patrilineal genealogies or from state generated household registers. Stevan Harrell, Ts'ui-jung Liu, Ted Telford, and Zhongwei Zhao and their associates use genealogies to estimate trends in demographic rates over the very long term, especially for southeastern, south central, and southern China.[1] We and Arthur Wolf and his and our collaborators use Qing dynastic (1640-1911) and Japanese colonial (1905-1945) household registers from Liaoning and Taiwan Provinces in northeast and southeast China respectively to examine associations between individual and household characteristics on the one hand and demographic behavior on the other, largely during the last two centuries.[2]

Several studies have identified shortcomings in these sources from internal evidence, comparison with predictions from demographic models, and micro-simulations. Lineage genealogies rarely record wives and daughters, and usually omit sons who die without male offspring, especially if they die during infancy or childhood in the remote past (Harrell 1987; Telford 1990). They rarely provide data on marriage timing and often do not even record any vital data at all (Harrell 2003). Household registers appear to record most wives, but not necessarily most daughters and may even miss some sons. While recording of vital events tends to be complete, some registers appear to under-record mortality especially among very elderly males (Lee and Campbell 1997). Most recently, micro-simulation revealed that the selectivity against extinct patrilines could distort estimates of demographic rates from genealogies, biasing mortality estimates downward and fertility estimates upward (Zhao 2001).[3] In a caveat, Zhao (2001) noted that other forms of selection could further bias estimates from genealogies, but could not be accounted for in his simulation model.

---

[1] See Harrell (2003); Harrell and Pullum (1995); Holman, Han, Harrell (2003); Liu (1978, 1982, 1985, 1992); Telford (1992, 1994); and Zhao (1997).

[2] See Campbell and Lee (1996, 2000, 2001, 2002, 2003, 2004); Campbell, Lee, and Elliott (2002); Chuang and Wolf (1995); Lee and Campbell (1997, 1998a, 1998b); and Wolf and Huang (1980).

[3] Zhao (2001) noted that these distortions could be eliminated by excluding the first few generations of the patriline from calculations, but only if the sole source of bias was the underrepresentation of extinct patrilines.

By direct comparison of records for the same population from lineage genealogies and from household registers, this paper demonstrates that lineage genealogies and related sources have inconsistencies not identified or accounted for in previous analyses. In particular, omission from lineage genealogies is by no means random. The chance that an individual listed in a household register is also recorded in their family genealogy depends on their life span, their reproduction, and their socioeconomic attainment. As a result of this selectivity, the demographic and social outcomes of individuals recorded in lineage genealogies differ systematically from the outcomes of their unrecorded kin.

Our paper is organized into three parts. We begin by introducing the household registers and lineage genealogies that we compare, and summarize our procedures to link these two sources. In part two, we apply multivariate techniques to identify the characteristics of individuals recorded in both the household registers and the lineage genealogies. In part three, we compare demographic and social outcomes to assess the possible bias associated with calculations that rely on lineage genealogies. To assess whether the descent groups for which genealogies exist are distinguished not only by their survivorship but other social and economic characteristics that Zhao (2001) suggested could create additional biases in estimates, we compare outcomes for descent groups according to whether or not a genealogy is available. To assess the effects of the demographic, social, and economic selectivity of recording within descent groups, we then compare outcomes for individuals according to whether or not they are recorded in their descent group's genealogy. In the conclusion, we discuss the implications of these findings, focusing briefly on fertility, and outlining the potential impact of the inconsistencies we identify on observed fertility levels and patterns.

The results in this paper are of interest not only to those who work on Chinese population and society, but also to historical demographers and sociologists in general. Until now, most micro-level quantitative analyses of society or population in the past have relied on single sources, or linked but highly complementary sources normally produced by similar institutions. This is one of a relatively small number of historical studies to make use of multiple totally independent sources that record the same population, and search for contradictions between the sources.[4] A further distinguishing feature of this study is that we compare official and privately produced sources. Thus we contrast official and private constructions of kinship and population in the Chinese past.

In our future work, we will turn this analysis on its head and use lineage genealogies to examine omissions from household registers. Our preliminary examinations have revealed that just as there were individuals in the household registers that were not listed in their lineage genealogies, so there were individuals in the lineage genealogies who were not listed in the household registers. Many, indeed probably most

[4] The only similar study of which we are aware is the one of African-American mortality by Sam Preston, Irma Elo and their collaborators at the University of Pennsylvania, which used matched records on individuals in a variety of different sources to better understand problems with age misreporting and their influence on estimates of African-American mortality (Elo and Preston 1994; Elo, Preston, Rosenwaike, and Cheney 1996; Hill, Preston, Elo, and Rosenwaike 1997; Hussey and Elo 1997; Preston, Elo, Foster, and Fu 1998; Preston, Elo, Rosenwaike, and Hill 1996; Preston, Elo, and Stewart 1999; Rosenwaike, Hill, Preston, and Elo 1998).

of these individuals were descendants of ancestors who escaped population registration and were classified in the household registers as *taoding*, literally "escapees." A systematic examination of the determinants of omission from the household registers, however, awaits further progress in organizing and linking the records from the lineage genealogies and generating the variables required for analysis.

I. Data

Our data comprises one of the larger and certainly longer individual level longitudinal panel data sets assembled for micro-level historical studies. All together we have linked as many as seventeen generations from the seventeenth century to the present with 275,000 individual histories, their households, their descent groups, and their demographic and social outcomes. The core of our historical data come from triennial household registers for almost 500 villages from Liaoning province dating from 1749 to 1909 which we have entered into machine readable form and linked to both other historical populations recorded in family genealogies and grave inscriptions from these same villages, and to other contemporary populations of their descendants recorded in current censal and household registers as well as retrospective surveys. We have also located and linked a variety of contextual information about the region and specific communities.

The linked data have six distinct features that make them uniquely suited to address a variety of substantive questions in historical demography and family sociology. First, they are longitudinal and individual-level and include not only demographic information, but social, economic, and political information as well. Second, they locate individuals within their households and kin groups, distinguishing kin by relationship and co-residence. Third, they include parallel information at the community level on local practices, economic conditions, and state policies. Fourth, they follow the population from their origins in the seventeenth century to the present. Fifth, they are numerous and varied enough to test many of the assertions about the relationships between kinship and demography over space and time. Sixth, they come from disparate but complementary sources to provide population information from the points of view of the state as well as the local population themselves.

We have been able to produce such historical data because of the internal consistency of the core household register data, their availability through the Genealogical Society of Utah and the Liaoning Provincial Archives, and the sustained efforts of teams of colleagues and data entry operators in the Peoples Republic of China. We have described the data and data entry operation elsewhere (Ding, Guo, Lee, and Campbell 2004; Lee and Campbell 1997; Lee and Wang 1999). In addition, since 1998 an on-going collaborative project with the Liaoning Provincial Local History Office allows us to visit these villages to collect historical and contemporary population sources, survey specific lineages, and record analogous contemporary information to the historical records. All together we have spent over 500 person-days in fieldwork visiting almost 50 of the largest villages to collect over 30 bound genealogies and over 50 genealogical charts and lists. We have also collected and transcribed dozens of long historical grave

inscriptions, half a dozen other inscriptions, and half a dozen contemporary village census or household registers. Most importantly we have completed retrospective and contemporary surveys in over a dozen villages recording each individual born in the village since 1949, their birth, marriage, death dates, education, occupation, and migration history and have linked these contemporary and historical populations.

Table 1 summarizes the currently linked data: 1.3 million observations of 225 thousand individuals who lived between 1750-1909 of which 1,066,004 observations for 187,389 individuals have been checked and cleaned, 80 largely patrilineal genealogies with some 25,000 largely male descendents and their spouses who lived between 1650-2000, 30 inscriptions from 1770-1940 with as many as 1000 linked relatives, and 11 retrospective surveys and 3 contemporary household registers with over 15,000 individuals born between 1880-2002. By supplementing the household registers with genealogies and other historical sources, we can trace 20,000 individuals from the arrival of their descent group founders in Liaoning in the late seventeenth century forward to the present. In addition, by surveying contemporary descendants from these historical populations and linking them to the registers, we can trace 50,000 people from the present back to the mid-eighteenth century.

Table 1. Demographic Sources for Northeast China

Population	Period	Registers	Observations	Coding	Genealogy	Inscription	Survey
Aerjishan	1813-1909	18	13,622	Done			
Bakeshu	1759-1909	32	48,709	Done	7	5	2
Changzhaizi	1768-1909	25	46,810	Done	2		
Chengnei	1798-1909	24	55,671	Done			
Dadianzi	1756-1909	27	76,984	Done	3	3	1
Dami	1759-1909	32	31,544	Done	2		
Daoyitun	1774-1909	35	118,633	Done	8	7	2
Daxintun	1749-1909	29	86,956	Done	10		1
Diaopitun	1768-1909	26	70,153	Done			
Feicheng	1756-1909	39	70,175	Done	8	5	
Gaizhou Manhan	1753-1909	20	50,110	Done			
Gaizhou Mianding	1789-1909	25	56,051	Done			
Gaizhou	1762-1909	27	42,834	Done	4		
Guosantun	1774-1909	34	35,073	Done	4	2	1
Haizhou	1759-1909	26	100000	Correcting			2
Langjiabao	1756-1909	25	47,340	Done			
Nianmadahaizhai	1749-1909	29	53,882	Done	4	9	1
Niuzhuang Liuerbao	1780-1906	23	50,253	Done			
Wangduoluoshu				Collecting			
Wangzhihuitun	1765-1909	28	60,339	Done		5	1
Waziyu	1777-1906	21	50000	Entering			
Wuhu				Collecting			
Zhaohuatun	1774-1909	26	50,865	Done	1	1	

Historical Household Registers

The historical household registers provide detailed information on social

outcomes, demographic behavior, and kinship organization for a population of hereditary royal peasants between 1749 and 1909. As summarized in table 1, we have completed data entry and data cleaning of the household registers for 19 administrative populations, are cleaning the data entry of a twentieth population, Haizhou, are entering a twenty-first population, Waziyu, and are collecting the registers for two more populations, Wangduoluoshu and Wuhu. In addition, we have launched a similar data project for 120 other villages in another northeastern Chinese province, Jilin.

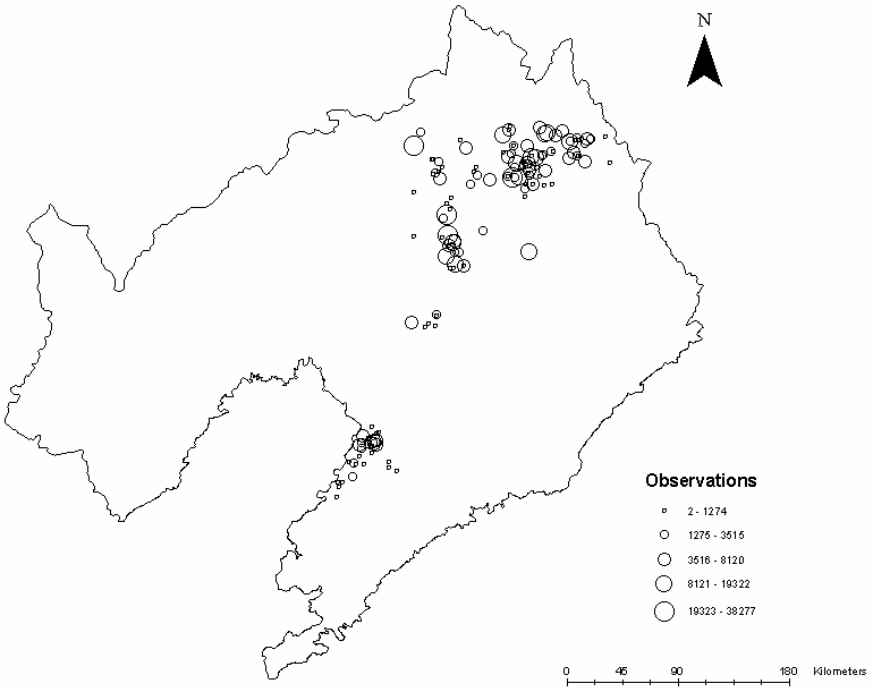
The institutional contexts of these populations varied dramatically. While most of these populations produced grain, several of them produced more specialized goods. The Dami population gathered honey, the Gaizhou Mianding population raised cotton, and the Diaopitun population produced animal furs. While most of these populations consisted of royal peasants, some such as Aerjishan were royal serfs. Others such as Gaizhou Mianding were in-between. As a result of such institutional variation, the opportunities for economic, educational, political, and social advancement varied across populations. Members of some populations were eligible to take state examinations, serve in state offices, and to earn state titles; others were not.

The registers record these populations more completely than almost any other historic rural population in China because they were affiliated with the imperial household as royal peasants or royal serfs, and because they were organized under the Han Martial Banners, and therefore liable for military service. The Imperial Household Agency surveyed and registered the population triennially beginning in 1749 with the establishment of the General Office of the Three Banner Commandry and designed a system of internal cross-checks to ensure data consistency and accuracy. First, they assigned every person in the banner population to a residential household called a *linghu* and registered them on a household certificate. Then they organized households into local household groups called *zu*, and compiled annually updated local household registers. Finally, every three years they compared these local registers and household certificates with the previous larger population and household register to compile a new register. They deleted and added people who had exited or entered in the last three years and updated the ages, relationships, and official positions of those people who remained as well as any changes in their given names. Each register, in other words, completely superseded its predecessor.

The registers list each individual one to a column in order of their relationship to the household head, with his children and grandchildren listed first, followed by coresident siblings and their descendants, and uncles, aunts, and cousins. Wives are always listed immediately after their husbands, unless a co-resident widowed mother-in-law supercedes them. For each person in the target population the registers report the following information: relationship to their household head; name(s) and name changes; adult banner status; age; animal birth year; lunar birth month, birth day, and birth hour; marriage, death, or emigration, if any during the intercensal period; physical disabilities, if any and if the person is an adult male; name of their household group head; banner affiliation; and village of residence. For adult males, the registers also record official titles and occupations that allow us to measure individual income or wealth. 4 percent of

males held such titles at some point in their life; they and their families comprise the rural local elite. For working-age males, the registers also record whether or not they were considered disabled. 10.2 percent of males were classified as disabled. Additional information, such as reproductive histories, are available through record linkage and comparison. Since individuals are listed in the same order in successive registers, longitudinal linkage of entries is straightforward.

As Map 1 shows, the 500 Liaoning villages are arranged in three distinct regions over an area of 40,000 square kilometers, approximately the size of the Netherlands: a commercialized south from Haicheng to Dalian down the Liaodong peninsula, an administrative center located on the Liaodong Plain around the provincial capital, and an agricultural north in the hills and mountain ranges directly north. These pronounced regional differences enable us to test a variety of hypotheses about socioeconomic conditions and demographic behavior, and measure regional characteristics as well as shared processes and relationships. The common immigration origins and institutional background of our communities allow us to control for such particular circumstances. While our results only illuminate the behavior of specific Chinese populations, we can draw from them implications for the demography not of China as a whole, but of specific social, economic, and political systems. This strategy, comparing local rather than national contexts, avoids the problem of representativeness normally inherent in community studies.



Map 1 Liaoning Historical Study Populations

These registers have a number of features that distinguish them as a source for historical demography. In contrast with historical Chinese demographic sources such as

genealogies that only record adult males, the historical registers record most boys and some girls from childhood, as well as all women from the time of their marriage. Unlike genealogies, they also provide detail on village and household residence. In contrast with parish registers, an important source for European historical demography, they allow for precise measurement of the population at risk of experiencing most demographic events and social outcomes. We have already used the registers to investigate the determinants of individual survivorship (Campbell and Lee 1996, 2000, 2002, 2004), marriage (Lee and Campbell 1998a, 1998b), migration (Campbell and Lee 2001), ethnic identity (Campbell, Lee, and Elliott 2002), and social mobility (Campbell and Lee 2003). These publications also detail the strengths and limitations of the register data relevant to the analysis of each outcome.

One of the most important features of these linked data is that they follow families for as many as seven generations, from the middle of the eighteenth century to the beginning of the twentieth. The population is closed, in the sense that the registers followed families that moved from one village to another within the region. Entries into and exits from the region were rare, and when they did occur, their timing was recorded (Lee and Campbell 1997, 223-237; Lee and Wang 1999, 149-153). Through linkage within the registers, therefore, we can identify the paternal kin of individuals, even if they live in other households or even villages. Table 2 summarizes the results of the linkage we have already carried out within the household registers. We can locate a great-great-grandfather within the registers for 50.2 percent of men overall, and 83.0 percent of men who first appear after 1900. Figure 1 presents this information in graphical form, presenting the proportions of children in each register for whom specified paternal ancestors can be located. The proportions are higher than in Table 2 because of the restriction to children.

Table 2. Males by Number of Generations of Ancestry in Registers (March 2004)

Paternal Ancestor	Percentage of males for whom specified ancestor can be located: All males	Appearing after 1900
Father	89.6	92.8
Grandfather	78.6	89.2
Great-grandfather	65.2	87.1
Great-great-grandfather	50.2	83.0
Great-great-great-grandfather	34.3	73.2
Great-great-great-great-grandfather	19.4	51.3
Great-great-great-great-great-grandfather	8.7	25.0
Great-great-great-great-great-great-grandfather	3.3	9.8
N	103402	23112

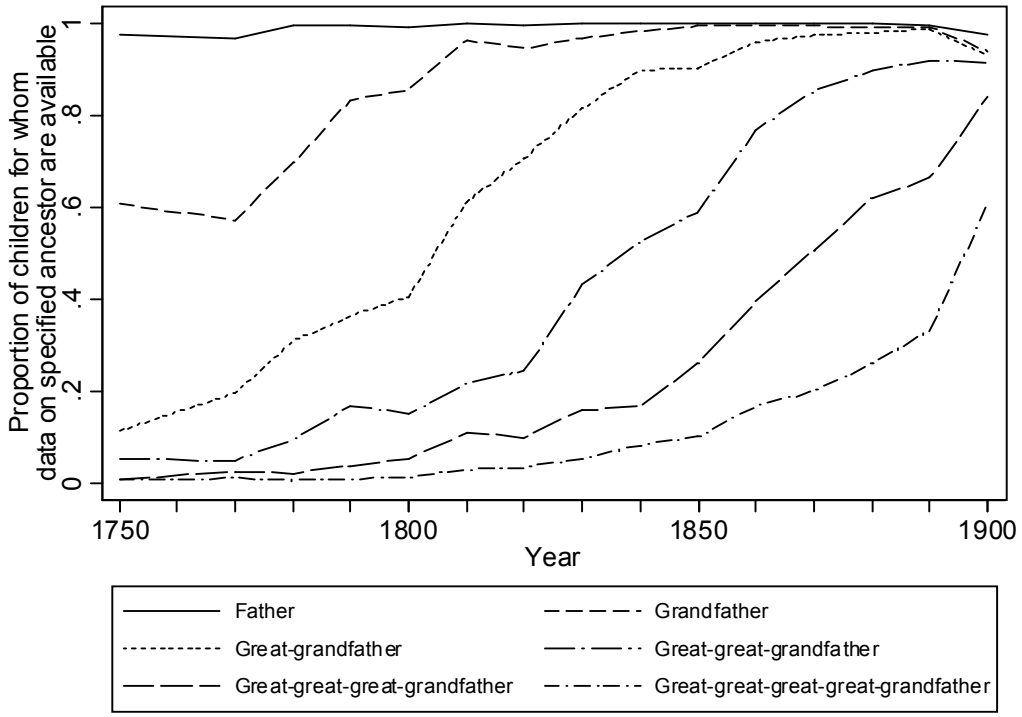


Figure 1 Children by numbers of generations that their ancestry can be traced in the registers

Through such linkage, we have grouped the individuals in the registers into descent lines and descent groups. First, the 187,705 distinct individuals described by the 1,066,004 observations in the registers are grouped into 16,901 descent lines defined by common descent from a paternal ancestor recorded in one of the registers. 7,562 of these account for 95 percent of the population. By assuming that household groups with the same surname who are listed consecutively in a register are related, these descent lines can be further aggregated into 1,920 descent groups defined by descent from a male founder who preceded the registers. The 758 largest groups account for 95 percent of the population. The small descent groups with only a few identified members tend to consist of the members of households that were first recorded in the registers in the late nineteenth century or beginning of the twentieth and could not be linked to a larger group.

Genealogies

Our genealogical data closely resemble those used in other analyses of Chinese historical demography. Consisting largely of lists of patrilineal ancestors arranged by generation, most of our genealogies are collections of names, organized by nuclear family and sub-descent groups, tracing the patrilineal population from the initial immigrant founder to the present day, identifying wives, elite success, and village of residence and in some cases providing information on vital events, especially adoption. All together, we have collected 80 such contemporary genealogies for some 40 different descent groups with a total population of over 25,000 largely adult males and their wives

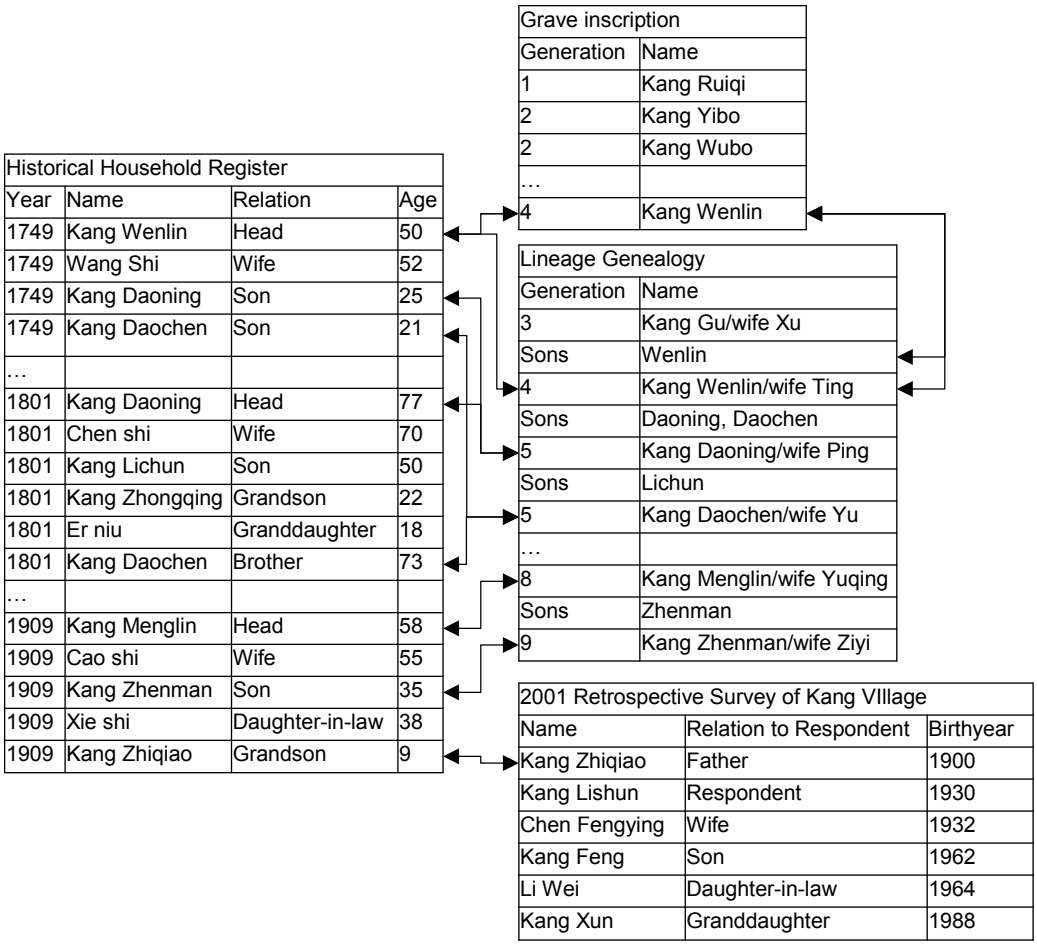
complemented by some 30 largely nineteenth century grave inscriptions which typically narrate individual and family histories on their front and back respectively as well as other local sources. While these genealogies have less vital information than the best European and American genealogical databases (Ewbank 2001), as the most current updates from a continuous tradition of current genealogies, they are far less retrospective and more complete. Moreover, what information they do miss can be found in complementary historical household registers and contemporary retrospective surveys.

Retrospective Surveys and Contemporary Household Registers

To date, we have collected contemporary household or censal registers or completed survey forms for over a dozen village descent groups with a total population of over 15,000 people. These typically not only record household residence, birthplace, and birth and marriage dates for contemporary residents but also education and occupational information for current residents as well as for all family members since 1949. To facilitate linking these data with archival data from the early twentieth century, the forms list the parents, grand parents, and great grand parents of the oldest member of each household as well as all their respective children noting dates of departure as well as death. Since these data cover the twentieth century, we do not consider them in this analysis.

Linkage

We have begun to link these sources with the historical household registers. Figure 1 provides a hypothetical example of a family recorded in the historical household registers and other sources, showing how the records of individual members may be linked. While virtually all the contemporary populations can be linked longitudinally to the household registers, this is less true of the historical populations recorded in the genealogies. As of the summer of 2003, we have linked more than 10,000 men from 40 of the more than 150 such sources that we have located with the household registers. These men account for nearly one-third of the men in the 40 sources. The remaining two-thirds are mostly men who lived before the registers began in 1749, after they ended in 1909, or in villages for whom we do not have household registers.



Historical Household Register			
Year	Name	Relation	Age
1749	Kang Wenlin	Head	50
1749	Wang Shi	Wife	52
1749	Kang Daoning	Son	25
1749	Kang Daochen	Son	21
…			
1801	Kang Daoning	Head	77
1801	Chen shi	Wife	70
1801	Kang Lichun	Son	50
1801	Kang Zhongqing	Grandson	22
1801	Er niu	Granddaughter	18
1801	Kang Daochen	Brother	73
…			
1909	Kang Menglin	Head	58
1909	Cao shi	Wife	55
1909	Kang Zhenman	Son	35
1909	Xie shi	Daughter-in-law	38
1909	Kang Zhiqiao	Grandson	9

Grave inscription	
Generation	Name
1	Kang Ruiqi
2	Kang Yibo
2	Kang Wubo
…	
4	Kang Wenlin

Lineage Genealogy	
Generation	Name
3	Kang Gu/wife Xu
Sons	Wenlin
4	Kang Wenlin/wife Ting
Sons	Daoning, Daochen
5	Kang Daoning/wife Ping
Sons	Lichun
5	Kang Daochen/wife Yu
…	
8	Kang Menglin/wife Yuqing
Sons	Zhenman
9	Kang Zhenman/wife Ziyi

2001 Retrospective Survey of Kang VIllage		
Name	Relation to Respondent	Birthyear
Kang Zhiqiao	Father	1900
Kang Lishun	Respondent	1930
Chen Fengying	Wife	1932
Kang Feng	Son	1962
Li Wei	Daughter-in-law	1964
Kang Xun	Granddaughter	1988

Figure 1 Hypothetical Example of Record Linkage

II. Determinants of Being Recorded

We begin with an examination of the determinants of being listed in a lineage genealogy or other source for individuals who appear in the household registers. We estimate logistic regressions where the outcome is an indicator of whether or not an individual was included in a genealogy. Right-hand side variables measure characteristics of the individual expected to affect the likelihood of being recorded. To test the assumption that men were more likely to be recorded if they lived to an advanced age, and had more descendants, we include the age at last appearance in the registers, their number of surviving sons at that time, and counts of their grandsons. To test the related assumption that the men in the descent group who had more close kin were more likely to be recorded, we also include a count of the number of male cousins. To test the hypothesis that more senior members of a generation were more likely to be recorded, we include seniority rank among first cousins. To assess the effects of socioeconomic attainment on the chances of being recorded, we include indicators of whether or not a man ever held different categories of official title or position in their lifetime. To examine whether more recent generations were better recorded, we include indicator variables to compare the periods 1749-1800, 1800-1850, and 1850-1909. To examine

how seniority within the kin group influenced the chances of being recorded, we include an indicator of their seniority among cousins.

We restrict the analysis to observations of men in descent groups for which a genealogy or other source has been collected. This ensures that comparisons are among the males who are actually at risk of inclusion in genealogy. Coefficients reflect effects of individual characteristics on the chances that a male in a descent group with a genealogy will be included in it, not their effects on the chances of being part of a descent group that has a genealogy. To ensure that coefficients do not reflect correlations between measured variables and unmeasured characteristics of the descent group or descent line, we also estimate a logistic regression with a fixed effect of descent line membership. The comparison in this fixed-effect model is between males who have a recorded male ancestor in common and are accordingly members of the same branch of the descent group. Coefficients are net of any effect associated with the tendency of some descent lines to be more likely to be recorded in a genealogy.

There were substantial differences between and within descent lines in the likelihood that members would be recorded in a genealogy or other source. Table 3 presents the results of the logistic regression. The substantial variation between descent lines, with some having many members recorded and some having none, is apparent from a comparison of the results of the two models. Moving from the first model to the fixed effect model, many of the coefficients increase in magnitude, and the overall 'fit' of the model increases, as reflected in the Pseudo-r^2. Individual characteristics, in other words, become better predictors of inclusion once the variation between descent lines is controlled for with a fixed effect. Further confirmation of differences between descent lines within the same descent group in the chances of being recorded is suggested by the finding in the fixed effect model that men who held civil positions were actually substantially less likely to be recorded in the genealogy than other men in the same descent line, even though they were not much less likely to be recorded than other members of their descent group.

Table 3. Logistic regression of inclusion in a genealogy, restricted to males in descent groups for which a genealogy is available

				With fixed effect of descent line		
	Mean	Odds ratio	p-value	Mean	Odds ratio	p-value
Proportion of males included	0.333			0.417		
Age at last appearance (Ref.: <= 20 *sui*)						
21-50 *sui*	0.225	1.153	0.05	0.228	1.135	0.02
51+ *sui*	0.341	2.080	0.09	0.326	2.838	0.00
Title or position ever held						
Banner	0.036	1.160	0.09	0.034	1.241	0.07
Civil	0.002	0.926	0.38	0.002	0.382	0.09

Exam	0.003	1.035	0.26	0.003	0.746	0.44
Purchased title	0.002	1.842	0.57	0.002	1.974	0.16
Honorific	0.002	1.568	0.74	0.002	4.315	0.05
Ever reported as disabled	0.125	1.328	0.07	0.133	0.946	0.44
Ever married	0.626	2.422	0.11	0.635	2.576	0.00
Number of living sons at last appearance	0.729	1.161	0.02	0.751	1.177	0.00
Number of grandsons ever born	5.183	0.999	0.00	5.282	1.004	0.04
Year of last appearance (Ref.: 1749-1799)						
1800-1849	0.197	2.226	0.16	0.177	2.937	0.00
1850-1909	0.684	5.059	0.34	0.727	4.553	0.00
Rank among cousins (1= eldest, 2 = next eldest etc.)	4.824	0.998	0.00	4.824	0.988	0.05
Number of cousins	8.610	0.992	0.00	8.610	1.007	0.16
Grandfather not identified	0.042	0.702	0.06	0.042	0.676	0.00
Observations			21692			13770
Log-likelihood			-12293.29			-6802.8
Pseudo-r^2			0.11			0.14

Even among men who are members of the same descent line, and accordingly relatively close kin, longevity determines whom the genealogy will record or omit. As expected from the past literature on the shortcomings of genealogies, longevity is important. According to the results in Table 3 for the model with a fixed effect of descent line, a man who survives past age 50 is 2.8 more times likely to appear in a genealogy or other source than a close relative who dies before age 20. A man who survived to between ages 20 and 15 was 13.5 percent times more likely to be recorded than someone who died earlier.

Reproductive success was similarly important: men with progeny were more likely to be recorded than men without. Every additional son still alive at the last appearance increased the chances of being recorded by one-sixth. Every additional grandson increased the chances of being recorded by another 0.4 percent. This apparently weak effect should be considered in the context of substantial variation in the number of grandsons. The standard deviation was 12.9. A one standard deviation increase in the number of grandsons increased the chances of being recorded by more than 5 percent. Men with more cousins were also more likely to be recorded. Every additional cousin raised the chances of being recorded by another 0.7 percent. Again, there was substantial variation in the numbers of cousins. The standard deviation was 7.45, thus a one standard deviation increase in the number of cousins raised the chances of being recorded by more than five percent.

Socioeconomic attainment is an important determinant of recording. Certain types of official titles or positions were more important. In the model with a fixed effect of descent line, men who had held a Banner position were twenty-four percent more

likely to be recorded, men who had held a purchased title were nearly twice as likely to be recorded, and men upon whom a honor had been bestowed were more than four times more likely. Civil positions, administrative positions that were not part of the Banners, actually lowered the chances of being recorded. Since the model controls for longevity and number of offspring, these effects are above and beyond any advantage in survivorship and reproduction associated with high socioeconomic attainment. Since the comparison in this model is between close relatives, this is also net of any tendency for descent lines that are especially successful at acquiring titles and positions to also be better recorded. Status within the family mattered as well: according to the model with a fixed effect, senior cousins were more likely to be recorded than junior ones.

Unsurprisingly, genealogies were most complete in recent decades. According to the model with the fixed effect of descent line, men who made their last appearance in the household registers between 1800 and 1850 were 2.9 times more likely to be recorded in a genealogy than men who made their last appearance in the eighteenth century, and men who made their last appearance after 1850 were 4.5 times more likely to be recorded. Genealogies were typically compiled retrospectively, with updates based on the recollections of surviving descent group members. Recent ancestors, of course, were more likely to be remembered accurately by the compilers of a genealogy than more distant ones.

III. Social and Demographic Outcomes of Descent Groups with Genealogies

We now assess the biases in genealogy-based estimates of demographic and social outcomes by examining how the descent groups for which genealogies and other sources are available differ from the population at large, and how members of descent groups who were recorded differed from those who were not. We estimate regressions for a variety of demographic and social outcomes in which the explanatory variables are membership in a descent group with a genealogy and individual inclusion in a genealogy. For mortality, we estimate separate logistic regressions for males and ever-married females where the outcome of interest is death within the next three years. For children, we estimate additional regressions according to whether or not the father was recorded in the genealogy. For ever-married females, we consider the effects on mortality of whether or not their husband was listed in a genealogy. For marriage, we also estimate two logistic regressions, one for whether or not a male has ever been married, and one for whether or not a male is currently married. For reproduction, we estimate Poisson regression for married males in which the outcomes are the numbers of male and female births in the next three years. Differentials in the chances of having, or at least registering, female births are of special interest because our previous work has shown that female survivorship was highly selective, and reflected family social and economic standing (Lee and Campbell 1997).

Finally, for socioeconomic attainment, we estimate logistic regressions in which the outcome is whether or not an individual holds different categories of official title or position. To control for geographic, age, and temporal variation in the likelihood of these outcomes, all estimations include appropriate controls for decade, five-year age group,

and register population. To examine socioeconomic aspiration, we also include an analysis of effects of recording a non-Han name in the registers. In our previous work we have shown that elite males in the population were more likely to switch a non-Han name, typically Manchu name, presumably as a reflection of their desire to align their ethnic identity with that of the rulers of the Qing, the Manchu.

For each outcome, we estimate two models. Table 4 presents the results. Model I assesses differences between descent groups according to whether or not they have a genealogy by including an indicator of whether or not an individual is a member of a descent group for which a genealogy is available. Estimated odds ratios compare outcomes for members of these descent groups, regardless of whether or not they as individuals are listed in the genealogy, to the population at large. Model II examines selectivity within the descent group by adding an indicator of whether or an individual was recorded in a genealogy. Odds ratios for this indicator reflect the effects of inclusion in the genealogy on the outcome of interest, above and beyond the effect of being a member of a descent group that has a genealogy. Table 4 includes an assessment of the net effect of being recorded in a genealogy, relative to the population at large, constructed by multiplying together the odds ratios for being the member of a descent group with a genealogy, and for being recorded in a genealogy.

Table 4. Regression of social and demographic outcomes on membership in a descent groups with a genealogy, and being recorded in a genealogy

			In a descent group with a genealogy (I)		Conditional on I, recorded in the genealogy (II)		Net effect of being recorded in a genealogy (I*II)
	N		Odds ratio	p-value	Odds ratio	p-value	
Mortality							
Never-married		I	0.929	0.44			
females (1-20 *sui*)		II	0.937	0.52	0.934	0.63	
Ever-married females	207473						
1-20 *sui*		I	1.030	0.79			
		II	1.088	0.46	0.771	0.22	0.839
21-50 *sui*		I	1.033	0.37			
		II	1.021	0.58	1.086	0.16	1.109
51-75 *sui*		I	1.097	0.01			
		II	1.069	0.08	1.182	0.00	1.264
Males	357998						
1-20 *sui* (according		I	1.005	0.90			
to own inclusion)		II	1.249	0.00	0.158	0.00	0.197
1-20 *sui* (according		I	0.938	0.21			
to father's inclusion)		II	0.926	0.15	1.090	0.27	1.003
21-50 *sui*		I	1.066	0.07			
		II	1.125	0.00	0.651	0.00	0.732
51-75 *sui*		I	1.104	0.00			
		II	1.066	0.06	0.910	0.01	0.970
Male ever-married	228100	I	1.215	0.00			
		II	1.204	0.00	1.067	0.00	1.285
Male currently married	228100	I	1.181	0.00			
		II	1.175	0.00	1.042	0.00	1.224
Marital fertility based on…							
Male births	125595	I	1.090	0.00			
		II	1.091	0.00	0.995	0.80	1.086
Female births	125595	I	1.232	0.00			
		II	1.236	0.00	0.978	0.71	1.209
Male possession of a non-	318188	I	1.162	0.00			
Han name		II	1.173	0.00	0.918	0.00	1.077
Male attainment of any	318188	I	1.696	0.00			
Official position or title		II	1.649	0.00	1.292	0.00	2.131
Banner position	318188	I	1.626	0.00			
		II	1.582	0.00	1.284	0.00	2.031
Exam	318188	I	1.698	0.00			

		II	1.632	0.00	1.436	0.00	2.344
Purchased	318188	I	3.168	0.00			
		II	3.011	0.00	1.723	0.00	5.188
Civil	318188	I	1.548	0.03			
		II	1.477	0.05	1.723	0.00	2.545
Honorific	318188	I	2.121	0.00			
		II	2.080	0.00	1.217	0.29	2.531

Note: Estimations also include controls for decade, five-year age group, and Banner population. To save space, we do not present these results.

Results for model I in table 4 confirm that the demographic and social outcomes of descent groups for which genealogies are available differ substantially from those of the population at large. Thus the descent groups for which genealogies are available differ not only in terms of their survivorship chances (Zhao 2001), but also in terms of socioeconomic and other characteristics. For most outcomes, membership in a descent group with a genealogy had beneficial effects. Men in descent groups that had genealogies or other materials available were also more likely to marry, more likely to register male and especially female births, more likely to record a non-Han name in the registers, and more likely to hold a Banner or civil position, more likely to have a purchased title, and more likely to have been awarded an honorific.[5]

The descent groups that had genealogies paid a price: their adult and elderly males actually suffered elevated death rates. This may reflect the 'price of privilege' noted in Lee and Campbell (1997) according to which many higher status males, including those who were heads of households or in some cases held official titles or positions, actually suffered higher death rates. Specifically, the males who were members of the descent groups for which genealogies are available may have been more likely to travel away from their villages, and accordingly more likely to contract disease. Elderly females in these descent groups were also more likely to die.

Reflecting the tendency apparent in previous studies and in table 3 for genealogies to be more likely to record men who lived longer, the death rates of the men recorded in genealogies were lower than those of other men in the same descent group. According to the results for model II, below age 20, their odds of dying were less than one-sixth, 0.158, those of other males in the same descent line who were not recorded in genealogies. Adult males recorded in a genealogy were one-third less likely to die than other members of the descent group, and elderly males were one-tenth less likely. The sons, daughters, and wives of the men recorded in genealogies, meanwhile, generally had unexceptional mortality, except that elderly wives had elevated mortality.

[5] The apparent contradiction between the reduction in chances of being recorded in a genealogy if a male holds a civil position in the fixed effect model in table 3, and the increase chances of holding a civil position if recorded in a genealogy in table 4, can be accounted for by recalling that the comparison in the fixed effect model in table 3 was between men in the same descent line, but in table 4 was between men in the same descent group.

When it came to marriage and reproduction, the advantage of being a member of a descent group that had a genealogy was more important than the advantage of being recorded in a genealogy. Thus according to the results for model II in table 4 men recorded in genealogies were slightly more likely than other descent group members to marry, but once married, their fertility was unexceptional. Thus according to table 4, men listed in genealogies were 4.2 percent more likely than other men in their descent group to be currently married, and 6.7 percent more likely to have ever married.

The net effect on demographic and social outcomes of inclusion in a genealogy, presented in the rightmost column of table 4, is large. The elderly wives or widows of men included in genealogies were more than one-quarter more likely to die than women in the population at large. Until age 20, men listed in genealogies had odds of dying that were one-fifth those of other men until age 20. As adults, their odds of dying were only three-fourths those of other males. They were also more likely than men in the population at large to marry, and once married, have children. They were also more likely to record a non-Han name and attain an official title.

Conclusion

This analysis reaffirms the value of comparing multiple linked demographic sources to identify and correct for the weaknesses of each (Elo and Preston 1994). At present, validation of sources used for historical demography typically relies either on checks of internal consistency, or comparison of patterns of demographic rates estimated from the source with those in other populations. Such approaches have successfully identified errors, omissions, and selectivity in a variety of sources, and have led to procedures for adjusting estimates to correct for these problems, or else suggested caveats to be kept in mind in the interpretation of results. Thus previous work has already identified 'holes' in Chinese genealogies (Harrell 1987), for which 'patches' in the form of adjustments or exclusions have been proposed (Telford 1990). The results here, however, suggest the possibility that there may be more subtle but pervasive patterns of error, omission, and selectivity that would not be apparent in a check of internal consistency or comparison with other populations.

Many of the specific findings confirm shortcomings of Chinese lineage genealogies suspected for some time, but indicate that problems in the selectivity with which members of a descent group were recorded were more serious than expected. For example, the finding that individual men who lived longer or attained higher socioeconomic status were more likely to be recorded hardly comes as a surprise. Analyses of male mortality in Chinese lineage genealogies typically only consider ages 20 and higher because the omission of men who died young is so common (Liu 1985, 1995). The finding that the death rates of males recorded in genealogies were lower those of other males between ages 20 and 50, however, challenges the widely used assumption that genealogies yield reliable estimates of male mortality above age 20. Similarly, while it has been suggested for some time that socioeconomic status may affect the chances of being recorded in a genealogy (Harrell 1987), the findings in Table 3 that descent group members who held an official title or position were more likely to be listed

in their genealogy, and in model II from Table 4 that men listed in a genealogy were more likely to hold an official title or position, raise the possibility that socioeconomic selectivity was more pronounced that previously thought.

The findings on fertility, meanwhile, identify a form of selectivity in recording that may actually bias fertility estimates from genealogies upwards. Previous estimates of historical Chinese fertility from lineage genealogies have tended to emphasize the downward bias in estimates induced by the omission from genealogies of men who died young (Telford 1995). Such analyses typically adjust for the omission of these births from the genealogy by multiplying fertility estimates by an adjustment factor based on assumed level of male mortality. The results here, however, confirm the existence of a countervailing bias that may lead fertility estimates from genealogies to overstate levels. Men with more children were clearly more likely to be included in a genealogy. According to Table 3, every additional son alive at the time of a man's last appearance raised the odds of being listed in a genealogy. As a result, according to Table 3, the odds men listed in genealogies were more likely to be married than other men, and once married, had higher fertility.

The results also make clear that there were two forms of selection that could affect estimates based on genealogies. First, as discussed above, genealogies were highly selective in terms of which descent group members they recorded. Second, perhaps even more importantly, the descent groups for which genealogies exist differ from other descent groups. While Zhao's (2001) simulation suggested that demographic selectivity, in the form of underrepresentation of extinct patrilines, might lead to underestimates of mortality or overestimates of fertility in calculations based on genealogies, it is clear from the results here that the descent groups for which genealogies have been located also differed socially and economically. Zhao's caveat (2001) that there could be other forms of selection in terms of which descent groups had genealogies, and which members were recorded, appears justified. Such social and economic selection actually may have been more important than demographic selection: even though on most social and economic outcomes descent groups with genealogies were more successful, their males actually experienced elevated mortality, the opposite of what would be expected if demographic selection was the only process at work.

Such unexpected and so far unexplained selectivity bias, in which the men in descent groups covered by genealogies actually had higher mortality, suggests that the individuals recorded in genealogies, and their families, may differ from the population at large in ways not previously suspected. Previous discussions of the implications of errors, omissions, and selectivity of genealogies have relied heavily on deduction, translating known characteristics such as the exclusion of males who died young into specific predictions about effects on estimates of rates. The results of our inductive exercise raise the possibility that there is an entirely separate class of problems associated with estimates from genealogies that can neither be predicted nor detected through checks of internal consistency or comparison with demographic models, because the selection processes that governed which descent groups had genealogies, and which members were recorded, were more complex than we realize.

This work is hardly complete. We are in still in the process of gathering and coding additional genealogies and related materials from Liaoning and linking them to the household registers. Thus we expect to replicate and refine the analysis here with progressively larger databases of materials from a wider variety of families in a more diverse set of villages. As these data accumulate, we will move beyond assessment of the overall levels of overestimation or underestimation, to examine variation among genealogies in the extent of omissions, errors, and selectivity. We will also address a variety of specific questions, for example, assessing the net bias in fertility estimates from genealogies by comparing the magnitudes of the upward bias induced by the omission of men who marry but have few or no children to the downward bias induced by the omission of men who died before they marry. Recognizing that the household registers are by no means perfect, we will also turn the tables in a future analysis, using the genealogies as a base from which to identify errors, omissions, and selectivity biases in the registers. We suspect that they are complementary and also asymmetric. Taken together, the view from above and the view from below may not produce the entire historical landscape.

References

Allen, Robert, Tommy Bengtsson, and Martin Dribe. Eds. Forthcoming. *New Evidence on the Standard of Living in Pre-industrial Europe and Asia*. Book manuscript accepted for publication by Oxford University Press.

Bengtsson, Tommy, and Osamu Saito. Ed. 2000. *Population and the Economy: From Hunger to Modern Economic Growth*. Oxford: Oxford University Press.

Bengtsson, Thomas, Cameron D. Campbell, and James Z. Lee, et al. 2004. *Life Under Pressure: Mortality and Living Standards in Europe and Asia, 1700-1900*. Cambridge, Ma.: MIT Press.

Campbell, Cameron, and James Z. Lee. 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: An International Quarterly* 1.3: 297-328.

_____2000. "Price fluctuations, family structure, and mortality in two rural Chinese populations: Household responses to economic stress in eighteenth and nineteenth century Liaoning." In Bengtsson and Saito, 371-420.

_____2001. "Free and unfree labor in Qing China: Emigration and escape among the bannermen of Northeast China, 1789-1909." *History of the Family: An International Quarterly* 6.4: 455-476.

_____2002. "When husbands and parents die: Widowhood and orphanhood in late imperial Liaoning, 1789-1909." In Derosas and Oris, 301-322.

_____2003. "Social mobility from a kinship perspective: Rural Liaoning, 1789-1909." *International Review of Social History*. 47:1-26.

_____2004. "Mortality and Household in Seven Liaodong Populations, 1749-1909." In Bengtsson, Tommy, Cameron Campbell, James Lee et al., 293-324.

_____forthcoming. "Living standards in Liaoning: Evidence from demographic outcomes." In Allen, Bengtsson, and Dribe.

Campbell, Cameron, James Z. Lee and Mark Elliott. 2002. "Identity construction and reconstruction: Naming and Manchu ethnicity in northeast China, 1749-1909." *Historical Methods*. 35.3 (Summer):101-116.

Derosas, Renzo, and Michel Oris. Ed. 2002. *When Dad Died: Individuals and Families Coping with Family Stress in Past Societies*. Bern: Peter Lang.

Ding Yizhuang. 1992. *Qingdai baqi zhufang zhidu yanjiu* (Research on the eight banner garrison system during the Qing). Tianjin: Guji chuban she.

Ding Yizhuang, Guo Songyi, James Lee, Cameron Campbell. 2003. *Liaodong baqi yimin shehui* (Liaodong banner immigration and rural society). Shanghai: Shanghai shehui kexue chubanshe.

Elliott, Mark. 2001. *The Manchu Way: The Eight Banners and Ethnic Identity in Late Imperial China*. Palo Alto: Stanford University Press.

Elo, I.T. and S.H. Preston. 1994. "Estimating African-American Mortality from Inaccurate Data." *Demography* 31:427-458.

Elo, I.T., S.H. Preston, I. Rosenwaike, M. Hill, and T. P. Cheney. 1996. "Consistency of Age Reporting on Death Certificates and Social Security Administration Records among Elderly African Americans." *Social Science Research* 25:292-307.

Ewbank, Douglas. 2001. "Demography in the age of genomics: a first look at the prospects." In Finch, Caleb, Wachter, Kenneth and Kevin Kinsella, Eds., *Cells and Surveys*, National Research Council, 64-109.

Fauve-Chamoux, Antoinette, and Emiko Ochiai, eds. 1998. *House and Stem Family in Eurasian Perspective*. Kyoto: International Research Center for Japanese Studies.

Hua Han. 2003. "Under the shadow of the collective good: An ethnographic analysis of fertility control in Xiaoshan, Zhejiang Province, China." Manuscript submitted to *Modern China*.

Harrell, Stevan. 1985. 'The rich get children: Segmentation, stratification, and population in three Chekiang lineages.' In *Family and Population in East Asian History*, eds., Susan B. Hanley and Arthur P. Wolf. Stanford: Stanford University Press.

_____1987. 'On the holes in Chinese genealogies.' *Late Imperial China* 8.2:53-79.

_____1995. ed. *Chinese Historical Microdemography.* Berkeley: University of California Press.

_____. 2003. "Figuring age at marriage from Chinese genealogies." In Melvin Thatcher Zhengmin Qian eds. *Chinese Genealogy and Local History Conference Papers*, Shanghai: Shanghai kexue jishu chubanshe, 90-110.

Hill, M., S.H. Preston, I.T. Elo, and I. Rosenwaike. 1997. "Age-linked Institutions and Age Reporting among Older African Americans." *Social Forces* 75: 1007-1030.

Holman, Darryl J., Han Hua, Stevan Harrell. 2003. "Birthspacing and early stopping behavior in Xiaoshan County, Zhejiang Province, China." Manuscript submitted to *Population and Development Review*.

Hussey, J. and I.T. Elo. 1997. "Cause-specific Mortality Among Older African Americans: Correlates and Consequences of Age Misreporting." *Social Biology* 44:227-246.

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

_____. 1998a "Getting a head: Headship succession and household division in three Chinese banner serf communities, 1789-1909." *Continuity and Change* 12.1 (May, 1998).

_____. 1998b. "Getting a head in Northeast China: Household Succession in four banner serf Populations, 1789-1909." In Fauve-Chamoux and Ochiai, 403-430.

Lee, James Z., and Wang Feng. 1999. *One Quarter of Humanity: Malthusian Mythology and Chinese Realities 1700-2000*. Cambridge, MA.: Harvard University Press.

Liu, Ts'ui-jung. 1978. "Chinese genealogies as a source for the study of historical demography." Studies and Essays in Commemoration of the Golden Jubilee of the Academia Sinica. Academia Sinica, 849-870.

_____1981. "The demographic dynamics of some clans in the lower Yangtze area, ca. 1400-1900." Academia Economica Papers 9.1 (January): 115-160.

_____1985. "The demography of two Chinese clans in Hsiao-shan, Chekiang, 1650-1850." In Hanley, Susan B., and Arthur P. Wolf, eds. 1985. *Family and Population in East Asian History.* Stanford: Stanford University Press, 13-61.

_____1992. *Ming Qing shiqi jiazu renkou yu shehui jingji bianqian* (Lineage population and socio-economic changes in the Ming and Qing periods). Taibei: Institute of Economics, Academia Sinica. 2 vols.

_____1995. A comparison of lineage populations in South China, ca. 1300-1900. In Harrell, Stevan (ed.) *Chinese Historical Microdemography*, Berkeley: University of California Press. 94-121.

Preston, S.H., I.T. Elo, Q. Stewart. 1999. "Effects of Age Misreporting on Mortality Estimates at Older Ages." *Population Studies* 53:165-177.

Preston, S.H., I.T. Elo, A. Foster, and H. Fu. 1998. "Reconstructing the Size of the African-American Population by Age and Sex: 1930-1990." *Demography* 35: 1-21.

Preston, S.H., I.T. Elo, I. Rosenwaike, and M. Hill. 1996. "African American Mortality at Older Ages: Results of a Matching Study." *Demography* 33:193-209.

Rosenwaike, I., M.E. Hill, S.H. Preston, and I.T. Elo. 1998. "Linking Death Certificates to Early Census Records: The African American Matched Records Sample." *Historical Methods* 31: 65-74.

Skinner, William G. 1987. 'Sichuan's population in the nineteenth century: Lessons from disaggregated data.' *Late Imperial China* 8.1 (June):1-79.

Telford, Ted A. 1986. 'Survey of social demographic data in Chinese genealogies.' *Late Imperial China* 7.2 (December):118-48.

_____1990. "Patching the Holes in Chinese Genealogies: Mortality in the Lineage Populations of Tongcheng County, 1300-1880." *Late Imperial China* 11.2 (December): 116-136.

_____1992. "Covariates of Mean's Age at First Marriage: The Historical Demography of Chinese Lineages." *Population Studies.* 46:19-35.

_____1994. 'Family and state in Qing China: Marriage in the Tongcheng lineages, 1650-1850.' *Jinshi jiazu yu zhengzhi bijiao lishi lunwen ji* (A Comparative History of the State and Family). Taipei: Academia Sinica, Institute of Modern History.

_____1995. 'Fertility and population growth in the lineages of Tongcheng County 1520-1661.' In Harrell 1995, 48-93.

Tong Yonggong, Guan Jialu. 1994. >Qianlongchao Shengjing zongguan neiwufu de sheli= (The establishment of the central office of the imperial household agency during the Qianlong period). *Gugong bowuyuan yuankan* 2: 19-23.

_____1999. >Shengjing neiwufu de shezhi yange jiqi lishi zuoyong= (The establishment of the Shengjing Imperial Household Agency and its historical functions). In Bo Di. Ed. *Qingzhu Wang Zhonghan jiaoshou bashiwu ji Wei Qingyuan jiaoshou qishi huayan xueshu lunwen heji* (Joint festschrift for Wang Zhonghan and Wei Qingyuan). Hefei: Huangshan shushe, 326-331.

Wolf, Arthur P., and Chieh-shan Huang. 1980. *Marriage and Adoption in China, 1845-1945.* Stanford: Stanford University Press.

Zhao Zhongwei. 1997a. "Long-term mortality patterns in Chinese history: Evidence from a recorded clan population." *Population Studies.* 51.2 (July): 117-128.

_____2001. "Chinese genealogies as a source for demographic research: A further assessment of their reliability and biases." *Population Studies*. 55.2 (July): 181-194.