Cameron Campbell
Lectures and Readings

Lecture 1: Kin networks, descent groups, and inequality in China

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

Campbell, Cameron and James Lee. 2008 *Social Science History*. 32(2):175-214.

Lecture 2: Comparative studies in quantitative history: The Eurasia Project and Beyond

Bengtsson, Tommy, Cameron Campbell, James Lee et al. 2004. Life Under Pressure: Mortality and Living Standards in Europe and Asia, 1700-1900. Cambridge: MIT Press. Chapters 1, 2, 3,e521Asia, 1700

Kin Networks, Marriage and Social Mobility in Late Imperial China

Cameron Campbell
UCLA
camcam@ucla.edu

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

2/24/2003

Prepared for presentation at the ISA RC28 meeting, Tokyo, March 1-3, 2003. Comments welcome. Please do not cite or circulate without contacting one of the authors for the latest version.

Social science has long assumed that kin networks influences demographic and social outcomes. This is especially true for China, where the high fertility of the East has long been assumed to be a product of a kinship system that encouraged early and universal marriage and redistributed resources to do so (Davis 1955; Malthus 1826/1986). Many historical social scientists have claimed to find patterns of demographic behavior consistent with such principles in China and indeed in all societies where complex extended families were common (Das Gupta 1997, 1998; Huang 1990; Skinner 1997; Wolf forthcoming). According to these scholars, resources were produced and shared collectively in particular by residential households, but among other kin as well. While the power vested in household heads by the state and in local descent group heads by customary rules meant that patriarchy and hierarchy were at the heart of collective production and consumption, custom also dictated that the prosperous assist less fortunate kin (Lang 1946, 181-189). These contrary tendencies towards protectionism and particularism therefore underlie current social theory about the relationship between domestic organization and demographic behavior in Eurasia in general and China in particular (Freedman 1958, 1966; Szonyi 2002; Zheng 2001).

This paper examines the influence of kinship on social and demographic outcomes in Liaoning Province in Northeast China during the late imperial period as an empirical test of these contradictory claims. We make use of one of the largest, longest, and most detailed panel data sets for an historical population: 161,000 individuals who lived in 500 village communities from 1749 to 1909, examining how kinship networks and household contexts influenced such social demographic outcomes as employment, marriage, and reproduction. Moreover, we contrast the pre 1860 Liaoning 'natural' economy with the post 1860 Liaoning 'treaty port' economy to test the common assertion and important assumption that kinship becomes less influential with the rise of commercialization, market penetration, and an increasingly open society.

China and Liaoning are especially appropriate places to study the influence of kinship on demographic behavior. Chinese kin groups not only influence demographic decisions, in many cases they actually make such decisions. Kin within and even without the household influence marriage, reproduction, education, employment, and even survivorship. Many Chinese kin groups had formal rules in imperial times to transmit family customs and strategies and to define the jurisdiction of kin authority by residence, family relationships, and gender (Ebrey 1984, 1991; Liu 1959).

Liaoning is one of the provinces where kin organization has been particularly well studied (Ding, Guo, Lee, and Campbell 2003). Liaoning is also ideal for a study of economic effects because of the previous simplicity of the Liaoning economy. As a frontier province, Liaoning only began to experience economic growth and subsequent commercialization with the arrival of the first settlers in the late seventeenth and early eighteenth century. In 1700, Liaoning was largely empty land. By 1930, Liaoning was already the most industrialized provincial economy. While the provincial population rose at the same time from several hundred thousand to several million people, a significant

proportion of these people farmed imperial estates and as royal peasants were unusually well documented.

Our research examines how kin proximity to people of power and property conditioned social and demographic outcomes. Specifically we study the internal organization of kin groups, and analyze how the numbers and presence or absence of specific kin, and numbers and presence or absence of specific kin with official positions, influenced demographic and social outcomes. We divide our paper into four parts. We begin in part one with some background on the subject and previous research. Then we turn in parts two and three to introduce the data and methods used in the analysis. Finally in part four we present our results.

Background

The Malthusian Paradigm remains influential in contemporary scholarship beginning with sociologists such as Davis (1948, 1955), historians such as Hajnal (1982), Laslett (1977, 1983, 1988), Macfarlane (1978, 1986, 1987, 1997), Schofield (1989), and Wrigley (1978), and most recently anthropologists such as Das Gupta (1997, 1998) and Skinner (1997). These scholars distinguish between two ideal model family systems: a relatively simple conjugal family system characteristic of Western, particularly northwestern Europe, and a comparatively more extended family system characteristic of a much wider geographic area stretching from East Asia and South Asia to Eastern and Southern Europe. Demographic historians have focused on describing the European conjugal family system and the preventive population check that characterized its demographic behavior. Their general conclusion is that while the social organization of such societies was relatively simple, their demography, and particularly their nuptiality, were sensitive to economic circumstances (Goldstone 1986; Levine 1987; Schofield 1985; Weir 1984; Wrigley and Schofield 1981). By contrast, the importance of kinship in the East shielded individual behavior from short-term economic fluctuations but rendered them vulnerable to social circumstances (Lee and Campbell 1997).

International comparisons of the influence of kin within the household on individual outcomes have confirmed the validity of such geographic comparisons, but have challenged our understanding of the links between kinship systems and demographic behavior. They have, for example, discovered little historical support for the long-held assertion that larger, more complex households better insulated members from economic pressure. Moreover they have not been able to substantiate many of the claimed behaviors above. Mortality rates from a comparison of eighteenth and nineteenth century rural communities were equally sensitive to short-term economic stress in southern Sweden, eastern Belgium, and northern Italy where households were relatively simple as in northeastern China and northeastern Japan where households were both larger and more complex. These same comparisons of mortality rates also demonstrate that widows, orphans, and motherless and fatherless children were actually more vulnerable to food price fluctuations in the joint Northeastern Chinese family than in the nuclear West European household (Bengtsson, Campbell, and Lee, et. al. 2003; Campbell and Lee 2002a).

A more complete understanding of the role of kinship systems in shaping demographic and social outcomes requires moving beyond the household to consider kin living elsewhere. The need for such analyses has long been recognized, but data limitations have hitherto precluded such research (Plakans 1984). Kin who lived apart interacted with each other in a variety of ways, sharing information as well as social, political, and economic resources. The genealogies that have been used in previous studies of kinship networks document kin ties, but do not provide information on residence, thus it is impossible to compare effects of kin according to whether or not they lived in the same household or village. Household registers document residence, but usually do not have adequate generational depth to reconstruct pedigrees and identify kin who lived outside the household.

This analysis is accordingly a substantial advance over previous efforts to study associations between kinship and social and demographic behavior. By longitudinally linking individuals for whom we have historical household registers over as many as seven generations, we can trace a subset of our population from the middle of the eighteenth century to the beginning of the twentieth, and reconstruct their kin networks. From 1789 onward, the registers organize individuals by household, thus we can identify which kin lived in the same household and which lived elsewhere, and compare their effects. In the future, with the additional collection of corollary auxiliary information on local economic, institutional, and social conditions we expect to relate behavior not just to kinship, but also to environmental circumstances, including economic circumstances and occupational history.

Data

The data we use are derived from ‘Household and Population Registers of the Eight Banner Han Army' (*Hanjun baqi rending hukou ce*). These household registers were compiled on a triennial basis for a number of Han banner populations in northeast China and certain other locations from the early eighteenth century until 1909. The Qing relied heavily on these registers for civilian and military administration of these populations. They accordingly devised a system of internal cross-checks to ensure consistency and accuracy. First, they assigned every person in the banner population to a residential household (*linghu*) and registered them on a household certificate (*menpai*). Then they organized households into clans (*zu*), and compiled annually updated clan genealogies (*zupu*). Finally, every three years they compared these genealogies and household certificates with the previous 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 banner registers provide far more comprehensive and accurate demographic and sociological data than the household registers and lineage genealogies common elsewhere in China (Harrell 1987, Jiang 1993, Skinner 1987, Telford 1990). This is

because the Northeast, which was the Qing homeland, was under special state jurisdiction, distinct from the provincial administration elsewhere. Regimentation of the population actually began as early as 1625, when the Manchus made Shenyang their capital and incorporated the surrounding communities into the banner system (Ding 1992, Elliott 2001). By 1752, with the establishment of the General Office of the Three Banner Commandry, not only was the population registered in remarkable precision and detail, migration was strictly controlled, not just between Northeast China and China Proper, but between communities within Northeast China as well. Government control over the population was tighter than in almost any other part of China (Tong and Guan 1994, 1999). Indeed, individuals who departed from the area without permission were actually identified in the registers as 'escapees' (*taoding*). As a result, the Eight Banner household registers are the most extensive and detailed records of a rural Chinese population in the late imperial period (Lee and Campbell 1997, 223-237).

The registers record at three year intervals for each person in the target population the following information in order of appearance: 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. Individuals are listed one to a column in order of their relationship to the 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 widowed mother-in-law supercedes them.

In additional to such social demographic data, the registers also record official positions. There are altogether five types of official positions: banner, civil service, examination, honorary, and household group leader. In our analysis of attainment we consider the first four of these categories. The first three are formal governmental offices and often included a salary and other perquisites. The fourth, honorary, were typically purchased and indicate personal resources or access to resources through the family. The fifth category, household group leader, or *zuzhang*, refers to the lowest level of local banner administration. We do not consider it in the analysis here. It was by far the most common position, with one for every few households, and did not include a salary.

The data we analyze here are from a sample of registers we have compiled that describes more than 100,000 individuals who lived in twenty separately registered populations in Liaoning province from the middle of the eighteenth century to the beginning of the twentieth century. Table 1 lists these populations and identifies the total number of available observations. Figure 1 summarizes the temporal distribution of the observations. The apparent increases in the numbers of available observation in the last half of the eighteenth century mostly reflects that relatively few registers from the middle of the eighteenth century survive, so that registers only become available in larger numbers at the end of the eighteenth century. The spectacular growth in the numbers of observations in the late nineteenth century reflects a combination of rapid natural increase in the population and the inclusion of new individuals or families in the register

population.

Table 1 and Figure 1 here

The registers are distinguished by the possibilities for record linkage across time and between kin. Individuals can be followed from one register to the next because they appear in almost the same order in successive registers. Accordingly, it is relatively straightforward to reconstruct life histories and generate variables describing such past characteristics as whether or not an individual had previously held official positions. Perhaps more importantly, by comparison of observations for the same individual in successive registers, we can construct outcome measures indicating whether or not particular events took place in the time interval between two successive registers. For this analysis, we construct indicators of whether or not men who without an official position attains one by the next register, whether or not men who have not yet married do so by the next register, and how many children a married man will father by the next register.

The extensive detail on household relationship, meanwhile, allows for reconstruction of genealogies and identification of kin living in the same or different households. Our basic procedure is to chain together the links between fathers and sons to identify grandfathers, great-grandfathers, and more distant male ancestors. Many of the men who appear in the later registers, for example, can have their ancestry traced back six or seven generations. Figure 2 summarizes time trends in the proportions of men for whom we identify fathers, grandfathers, and great-grandfathers. Once we have constructed genealogies, it is a straightforward matter of data processing to identify brothers, cousins, first cousins, second cousins, and other kin and measure their characteristics, regardless of whether they are in the same household or not. At present we can only do this for paternal kin, not maternal kin, because we have not yet traced the wives recorded in the registers back to their natal households.

Figure 2 here

The data have some additional limitations relevant to the analysis. First, they do not record any employment other than official employment. If any family members had occupations other than as employees of the state, there would be no record. If the commercialization of the late nineteenth century created new opportunities for employment outside the state bureaucracy, the registers do not record it. Second, the data do not record income or assets, thus it is impossible to consider effects of family landholding or wealth. Third, they fail to record children who died in the first few years of life, before they were old enough for their parents to register them. Outcome measures for an analysis of reproduction does not include these births, and is based solely on children who survived long enough to be registered. Differences in reproduction apparent in the analysis may reflect differences in both fertility and infant and early childhood mortality. Fourth, the registers may omit a very small number of marriages in which a woman joined her husband's family after one register and died before the next.

The requirements of the analysis and the limitations of the data allow us to make use of only a subset of these observations. First, we restrict to our analysis to males. Only males were eligible for official positions. An analysis of female first marriage was impractical because the registers omitted many daughters and recorded women only when they were wives in their husband's household. Second, we restrict to registers from 1789 or later years, because the earlier registers did not distinguish individuals by household. Third, our discrete-time event history approach limits us to registers for which the one immediately succeeding or the one after it were also available. Fourth, we only include observations of men for whom a father could be identified. For each analysis, of course, we apply additional restrictions, as described later in the section on methods.

Methods

To investigate how kin networks shaped social and demographic outcomes, we apply discrete-time event-history methods. For the analyses of attainment of position and first marriage, we estimate logistic regressions. The outcome measure in the analysis of the attainment of position is a dichotomous indicator or whether or not a man acquires a position by the next available register. We restrict the analysis to men who have not yet acquired a position. The outcome measure in the analysis of first marriage, meanwhile, is a dichotomous indicator of whether or not a man marries for the first time by the next available register. We restrict the analysis to men who have not yet married. For the study of reproduction, we use Poisson regression. The outcome measure is a count of the number of males recorded as born to the individual by the next available register. We restrict to observations of ever-married married men. In all of these analyses, we only use observations where either the immediately succeeding register or the one after it is available.

We examine attainment, marriage, and reproduction because of their sensitivity to allocations of economic, social, and political resources makes them ideal for reconstructing the internal dynamics of the kin group. Official positions were ostensibly awarded accorded to merit. The more prestigious and lucrative ones required skills that would have required investments in education. To the extent that the bureaucratic allocation of positions made the process vulnerable to particularism, families had to mobilize social and political resources to acquire them for specific members. Marriage, meanwhile, not only reflects a decision by the groom's family to allocate the resources for the acquisition of a spouse, it also reflects an explicit assessment on the part of the bride's family of the groom's kin group, and his standing within that group. Reproduction was also subject to the control of couples and the larger family (Lee and Wang 1999). Not only was fertility itself subject to control, but the chances that a child would survive long enough to appear in the registers used here depended on additional resource allocation by parents and children. We do not examine mortality here because our previous analyses have shown that its relationship with well being and access to resources was complex (Campbell and Lee 1996, Campbell and Lee 2000b). For example, possession of a position actually seems to have been associated with higher mortality for some males because the benefits associated with increased consumption

were more than offset by a higher risk of exposure to infection.

We compare four concentric circles of kin. The innermost circle comprises the father-son dyad. Next come brothers. After that come men who are also descended from the index individual's grandfather, that is, cousins and uncles. Finally we consider men who are descended from the index individual's great-grandfather. These include second cousins and father's cousins. Our expectation is that characteristics of more distant kin will be less important for outcomes. The precise pattern of effects according to distance, of course, will provide insight into kin group organization. While there are obvious reasons to expect father's characteristics to be very important, and brother's characteristics to be somewhat important, expectations for more distant kin are unclear. The ideology of solidarity within the larger kin group conflicted with the difficulties and even drawbacks of sustaining ties with distant relatives.

We focus on three aspects of the kin network: positions held by kin, numbers of kin, and individual seniority within the kin network. Table 2 summarizes the variables of substantive interest. The measures of positions held by kin are dichotomous, indicating whether or not the index individual has any kin of the specified type who holds a position. Comparison of effects of positions held by kin according to their proximity identifies the boundaries of the kin group and map flows of social, political, and economic resources.[1] Positive effects of having a relative with a position, for example, indicate that the relationship carried with it access to social, political, or economic resources. Adverse effects, meanwhile, reveal contention within the kin group. Lack of an effect, meanwhile, indicates that the specified relationship was not part of the kin group that determined the outcome of interest.

Table 2 here

We also examine the effects of numbers of kin. In nineteenth-century rural Liaoning, most people were not fortunate enough to have a relative who held a position. For such people, the most important feature of the kin network was its size. Larger kin networks had more options for sharing economic, social, and political resources, whether by cooperating in agricultural work, sharing information, personal connections, and economic resources, or taking advantage of their size and solidarity in disputes with other families. By comparing the effects of numbers of kin of different types, we identify the boundaries that constrained such interactions. Relatives whose numbers did not affect demographic and social outcomes were not part of the locus in which the decisions that affected such outcomes were made.

Comparisons of the effects of seniority among brothers, cousins, second cousins,

[1] For father and grandfather, the indicator measured whether or not they had ever held a position in their lifetime. For uncles, father's cousins in the same household, and father's cousin in the same household, the indicator measured whether or not any of the specified kin who were alive at the time the index individual was first observed had held a position by that time. For brothers, the indicator measured whether or not currently living brothers held a position or had held one in the past.

and household members identify the locus within which family members collectively set priorities. For example, to the extent that parents were largely responsible for decisions about the marriage of their sons, only seniority among brothers should have affected marriage chances. To the extent that marriages were decided on by the larger household, seniority within the household should have been more important than seniority among brothers. To the extent that the kin group beyond the household was the relevant locus, seniority among second cousins should have been important. Similarly, examination of the role of seniority in determining attainment chances identifies the locus within which decisions about the allocation of resources and use of connections that affected the chances of obtaining a position were made.

To assess the role of the household as a unit of organization distinct from the larger kin group, we compare the effects of characteristics of distant kin by whether or not they lived in the same household. To the extent that the interactions that governed demographic and social outcomes took place largely within households, and ties between kin living in separate households were weak, the characteristics of kin who lived outside the household should not have influenced these outcomes. Conversely, if the boundaries between households were porous, and kin who lived apart shared economic, social, or political resources, then the characteristics of kin beyond the household should have mattered.

To account for secular changes in attainment, marriage, and reproduction, we include an indicator of whether or not the individual concerned was born after 1840. Individuals born after 1840 spent their entire adulthood after in the period of increasing commercialization, rapidly rising population, and decreasing opportunities for attainment of official position that began around 1860.[2] Results from previous analyses suggest that overall, the last half of the nineteenth century was nevertheless a period of rising living standards (Campbell and Lee 2000a).[3] Trends in attainment, marriage chances, and fertility, summarized in Figures 3 through 5, are broadly but not perfectly consistent with this characterization. Because the population grew in size while the number of official positions remained constant and eventually fell, individual chances of obtaining a position in Figure 3 declined. According to Figure 4, marriage rates declined until the 1860s, then began rising. According to Figure 5, fertility peaked in the 1870s and 1880s. The chances that men would marry early increased, though the proportion of men who ever married remained stable. Reproduction increased, though given the limitations in the recording of children who died early noted earlier, this could also have reflected reductions in infant and early child mortality.

[2] We also estimated models that compared all observations of men after 1860 with those before. The results were more ambiguous, we believe because observations of men after 1860 include a substantial proportion of men who had the opportunity to marry or attain a position as adults before 1860, but had failed.

[3] For example, mortality and fertility became less sensitive to economic fluctuations, suggesting that families were no longer living as close to the margin as in the eighteenth and early nineteenth centuries (Campbell and Lee 2000a).

Figures 3 through 5 here

We also include a variety of control variables to ensure that coefficients do not reflect compositional differences between subpopulations. We account for age effects with dichotomous indicator variables for five-year age groups. We account for geographic variation with set of dichotomous indicator variables for each state farm population. We also include separate dichotomous indicator variables to identify the observations of men who could not be linked to their grandfathers or great-grandfathers and for whom the relevant measures of kin could not be constructed. For these observations, the affected kin measures are all set to zero.[4] Finally, in the analyses of attainment and marriage, we included indicators for whether or not the next available register was six years away.

For each of the three outcomes of interest, we estimate a basic model, a model with a fixed effect of kin group, and a model with cohort interactions. The first is a basic model that assumes independence among the observations, in the sense that related individuals do not share unobserved characteristics that affect both the outcomes of interest and explanatory variables. Such a model, while adequate to describe associations, cannot rule out the possibility that they reflect influence of such unobserved characteristics. For example, a positive effect of father's position on the chances of marriage might simply reflect a tendency for certain kin groups to be especially successful at obtaining both positions and spouses for their members.

To account for unobserved characteristics of kin groups that may affect both outcomes and explanatory variables, we estimate models in which we include a fixed effect of kin group and time. Specifically, we assume that at each point in time, men who have a great-grandfather in common share a higher or lower propensity for each of the outcomes as a result of their membership in a kin group. Estimated coefficients in this model reflect associations between outcomes and explanatory variables *among* members of the same kin group, net of differences between kin groups. For the examinations of attainment and marriage, we estimate a conditional logit, in which the underlying assumption is that one member of the kin group will experience the outcome of interest by the next register, and the coefficients reflect effects on chances of being that one member. Similarly, for the examination of fertility, we estimate a fixed effect Poisson regression.

To assess the implications for kinship of the changes that took place in the last half of the nineteenth century, we estimate a third model that includes interactions the indicator for birth after 1840 and the measures of kin network. For attainment, we examine whether the reduction in the chances of obtaining a position in the last half of the nineteenth led to an increase or a decrease in the role of the family characteristics in securing such positions. Reduced chances of attaining a position may have reduced the importance of family characteristics by increasing competition and increasing the relative importance of merit in the recruitment process. Conversely, reduced chances of

[4] We also estimated models restricting to observations of men for whom grandfathers and great-grandfather could be identified. Results for the relevant kin variables were the same.

attainment may have had the opposite effect, with increased competition giving a greater advantage to the families that already had position. For marriage and reproduction, we examine whether the earlier marriage and higher fertility of the last half of the nineteenth century was associated with a reduction in the importance of family background to differences between individuals.

Results

Kin influence on attainment, marriage, and reproduction varied by relationship as well as by residential arrangements. We conceive of kin networks as a series of concentric loci from close to increasingly remote relatives with decreasing interest in and influence on individual behavior. We therefore organize our discussion of the influence of kin on individual outcomes according to their proximity to each individual. We define the center of each individual's social world to be his relationship with his father, which is the father-son dyad.[5] Next closest were brothers, since sibling relationships differed from parental relationships, followed next by uncles and cousins, that is the kin connected to ego through his grandfather, followed by father's cousins and second cousins, that is the kin connected to ego through his great grandfather. These loci correspond roughly, but not exactly, to the first three of the traditional Chinese 'five degrees of mourning' which delineate mourning rituals and responsibilities (Feng 1937).

Tables 3-5 here

Fathers and Sons

According to our analyses, the father-son dyad was the most important locus for the determination of such outcomes as marriage and attainment in particular. Thus according to the analysis for Model I in Table 3, men whose fathers hold or held a position were 7.58 times more likely than other men to obtain a position by the next register. Results from Model II that included a fixed-effect for common great-grandfather underscore the importance of the patriline. Holding father's status constant, men whose grandfather had held a position were 31 percent more likely to attain one than members of their kin group whose grandfathers had not held a position.[6]

[5] See Hsiung 1994 for a vivid description of the influence of mothers on marriage, education, and other attainment.

[6] Turnover among the elites of Liaoning was nevertheless similar to that in the limited number of historical North American and European populations for which relevant studies have been carried out. A previous examination showed that only about one-third of the sons of men with position in Liaoning would attain positions of their own, and that between half and two-thirds of the men with position in each generation were 'new' in the sense that no one in their extended family held position (Campbell and Lee 2003, 19-20). In the European and North American populations for which results were available, between one-half and two-thirds of the sons of men in the highest occupational classes ended up in those classes themselves. Typically, one-third to one-half of the men in these classes were 'new' in the sense that their fathers had not been in the same occupational class.

The father-son dyad was also an important determinant of marriage chances. Father's and own possession of a position were both important determinants of marriage chances. According to results for Model I in table 4, father's and own possession of a position both increased the chances of marrying. Own position had the most powerful effect: men who held a position were 71 percent more likely to marry by the next register. Having a father who held a position had the next strongest effect, raising marriage chances by 44 percent. These effects all persisted in the face of an introduction of a fixed effect of kin group in Model II, confirming that the measures of position are not simply capturing the otherwise unobservable status of the larger kin group. Differences in the marriage chances between paternal cousins according to the possession of position by their fathers or selves were almost as pronounced as differences between unrelated men.

Surprisingly, however, father's and own position had little measurable influence on reproduction.[7] In particular, once we control for kin group membership, men who

unmarried brother was 20 percent more likely to marry than his younger, unmarried brothers. According to a calculation not shown here, the beneficial effects of seniority did not vary by whether or not the father was still alive, suggesting that this reflected decisions by brothers themselves or the larger kin group, not a preference exercised by the father.[8] That eldest surviving sons were so advantaged is hardly a surprise in light of their importance in traditional Chinese kinship.

Uncles and Cousins

Whereas relationships between fathers and sons as well as between brothers were characterized by solidarity, in the sense that outcomes were positively correlated, the picture for uncles and cousins hints at contention. While the Qing state appears to have in its allocation of positions for vertical transmission from fathers to sons, a contradictory desire to spread positions around led the competition among cousins to be zero-sum. Even though being the son of a man with a position improved attainment chances, being his nephew lowered them. Thus according to the results for Model I in table 3, the possession of a position by an uncle actually reduced the chances that his nephew would obtain one by about one-third.[9]

Cousins could nevertheless be of some benefit. According to Model I in Table 4, men with more cousins were more likely to marry. In particular, each additional cousin increased the chances of marrying by five percent. This was not because members of the kin groups that were more successful at securing brides and expanding through reproduction were more likely to have cousins. According to the results for Model II, differences in marriage chances between members of the same kin group according to the number of their cousins were as pronounced as those between unrelated men.

More Distant Kin

More distant kin still affected attainment chances, even when they lived in other households. According to Model I in Table 3, a man whose father's cousin held a position was about one-quarter to one-third more likely to acquire one by the next register. Whether or not the father's cousin with position actually lived in the same household was unimportant. Introduction of a fixed effect of kin group had little influence on the magnitudes of the effects, confirming that in a kin group in which a member of a senior generation held a position, the most advantaged members of the next generation were his sons, followed by his cousins' sons, followed by his unfortunate nephews.

[8] For attainment, in a version of Model I that included an indicator variable for the presence of the father and an interaction between it and the indicator for being eldest brother, the odds ratio for the interaction term was 0.84, with a p-value of 0.25. The direct effect of present of father was strong, with an odds ratio of 1.53 and a p-value of 0.002. For marriage, the odds ratio for the interaction was 1.04, with a p-value of

[9] Introduction of a fixed effect of kin group in Model II leaves the magnitude of this effect unchanged, confirming that it reflected pronounced differences within kin groups, and was not an artifact of differences between them.

Effects on marriage chances of the characteristics of more distant kin, however, depended on whether or not they lived in the same household. According to the results from Model I in table 4, a father's cousin who held position increased the chances of marriage by 17 percent if he lived in the same household. Men with more second cousins were also more likely to marry. Each second cousin raised the chances of marrying by 5 percent. When these distant kin lived outside the household, however, effects of their characteristics were very different. A father's cousin who held a position and lived outside the household lowered marriage chances. Additional second cousins living outside the household had no effects on marriage chances.

Seniority was important as well, but the relevant kin group differed for attainment and marriage. For attainment, seniority among kin in the same and other households was important. According to Model I in Table 3, the eldest male among a set of paternal second cousins was 1.25 times more likely to obtain a position by the next register than his younger relatives. Seniority among males in the household was relatively unimportant for attainment, especially after the inclusion of a fixed effect of the kin group in Model II. For marriage, seniority within the household was much more important than seniority in the larger kin group. According to Table 4, the eldest never-married male in the household was 70 percent more likely to marry by the next register than his younger never-married kin. The eldest never-married male in a kin group, however, had no advantage over his younger cousins and second cousins after a fixed effect of having a common great-grandfather was introduced.

Secular change

In spite of the economic and other changes that took place after 1860, kin networks actually became more important for attainment. Family background, in particular fathers' and brothers' position, became much more important for obtaining the official positions that were available. Results from model III in table 3 indicate that for men born after 1840, the advantage associated with having a father who held a position nearly doubled. For men born before 1840, having a father who held a position multiplied the chances of obtaining one by 6.52. For men born after 1840 it multiplied the chances of obtaining one by 12.13. The advantage associated with having a brother who held a position also seems to have increased, by a factor of about 1.5. The increase, however, is not statistically significant except by a very liberal criterion.

Conversely, kin networks seem to have become less important for marriage. According to Model III in Table 4, the advantage associated with having a father who held a position declined by about one-quarter. The benefits associated with additional brothers also declined somewhat. Similarly, disadvantages associated with having a grandfather or uncle who held position that were apparent for men born before 1840 were less pronounced for men born afterward.

Conclusion

In Qing Liaoning, kin networks beyond the nuclear family influenced the demographic and social outcomes of their members. In this analysis, we have demonstrated that the configuration of the kin network around the individual affected their chances of attaining official position and marrying. First, senior kin mattered. As was the case in almost all societies for which studies have been carried out, parental characteristics affected attainment outcomes. By taking advantage of the possibilities for record linkage and identification of distant kin, we have also shown that positions held by other senior kin influenced attainment and marriage chances, and that numbers of distant kin of the same generation influenced marriage chances.

Apparently, most sharing of the political, social, economic or other resources needed to marry or acquire a job appears to have been 'vertical' or 'horizontal.' Characteristics of members of the patriline such as the father and grandfather were important, as were characteristics of members of the same generation, including brothers, cousins, and second cousins. 'Diagonally' related kin appear to have been less important. Father's cousins were less important than fathers, though positions held by them did positively affect attainment and marriage chances. Positions held by uncles actually reduced attainment chances, and had no effect on marriage chances.

These results also begin to delineate the different roles played by the household and the larger kin group in shaping social and demographic outcomes. For attainment, social and political resources available through the larger kin group were more important. Positions held by father's cousins improved attainment chances, even if they lived in another household. Seniority among second cousins was a more important determinant of attainment than seniority within the household. The situation for marriage was reversed. The social, political, and economic resources available through the household appear to have been more important. Thus positions held by father's cousins were only beneficial if they lived in the same household. Seniority among the unmarried males within the household was far more important than seniority among second cousins.

The effects we observe, moreover, are clearly not artifacts of a tendency for some kin groups to be more successful than others at acquiring both positions and spouses for their members. In an analysis that failed to account for unobservable characteristics of kin groups, such as their status in local society, their wealth, or conditions in the village in which they lived, apparent effects of characteristics of specific kin on demographic and social outcomes might simply reflect the tendency of all the members of better-off kin groups to share an increased propensity to attain a position or marry. By estimating models that included a fixed effect of the kin group and thereby accounted for unobservable characteristics that its members had in common, we ensure that effects reflect differences within kin groups, not between them. In the case of attainment and marriage, effects of characteristics of specific kin almost all persisted, reflecting the importance of location within the kin network. In the case of reproduction, effects of kin largely disappeared, suggesting that measured associations in the model without a fixed stemmed from the tendency of members of better-off kin groups to all have elevated fertility.

While our work is by no means done, such findings demonstrate the potential for the use of quantitative approaches in to investigate a topic that has been previously been amenable only to qualitative approaches. While the data have been able for some time to allow systematic investigation of the influence of characteristics of close kin on demographic and social outcomes, until now assessments of the organization and implications of larger kin network beyond the nuclear family have relied almost exclusively on qualitative evidence. As a result, discussions of the role of the larger kin network in shaping individual outcomes have relied heavily on deduction, not induction. Assumed properties of the extended family are treated as first principles and predictions for demographic and social outcomes derived, for example, in Skinner (1997). Through analyses like the ones here, we intend to test the claims about the properties and implications of the extended family that have accumulated in the literature.

We expect the view of the kin network that emerges to be much more nuanced than would be expected from the existing literature. Rather than there being one identifiable kin group with fixed boundaries that affected outcomes, the work results here suggest that the importance of particular kin varied according to the outcome under consideration. For some outcomes, the nuclear family may have predominated. For others, for example marriage, the household appeared to the most important actor. For still others, the larger kin group was important. For fertility, kin group membership mattered, but position within the kin group appeared not to.

REFERENCES

Bengtsson, Thomas, Cameron D. Campbell, and James Z. Lee, et al. 2003. *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.

_____2000a. "Living Standards in Nineteenth Century Liaoning: Evidence from Demographic Outcomes." Presented at the conference "New Evidence on the Standard of Living in Preindustrial Europe and Asia," August 1-5, 2000, Arild, Sweden. Accepted for inclusion in edited conference volume under review at Oxford University Press.

_____2000b. "Price fluctuations, family structure, and mortality in two rural Chinese populations: Household responses to economic stress in eighteenth and nineteenth century Liaoning." In Bengtsson, Tommy, and Osamu Saito, Eds. 2000. *Population and the Economy: From Hunger to Modern Economic Growth*. Oxford: Oxford University Press, 371-420.

_____2002a. "When husbands and parents die: Widowhood and orphanhood in late imperial Liaoning, 1789-1909." In Derosas, Renzo, and Michel Oris, Eds. 2002. *When Dad Died: Individuals and Families Coping with Family Stress in Past Societies*. Bern: Peter Lang, 301-322.

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

Das Gupta, Monica. 1997. "Kinship systems and demographic regimes." In Kertzer and Fricke, 36-52.

_____1998. "Lifeboat versus corporate ethic: Social and demographical implications of stem and joint families." In Fauve-Chamoux and Ochiai, 444-466.

Davis, Kingsley. 1948. *Human Society*. New York: Macmillan.

_____1955. "Institutional factors favoring high fertility in underdeveloped areas." *Eugenics Quarterly* 2: 33-39.

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

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

Ebrey, Patricia. 1984 *Family and Property in Sung China: Yuan Ts'ai's Precepts for Social Life*. Translated, with annotation and 171 page introduction. Princeton: Princeton University Press.

_____1991 *Confucianism and Family Rituals in lmperial China: A Social History of Writing About Rites.* Princeton: Princeton University Press.

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

Feng Han-chi. 1937. 'The Chinese kinship system.' *Harvard Journal of Asiatic Studies* 2:142-289.

Freedman, Maurice. 1958. *Lineage Organization in Southeastern China.* London: Athlone Press.

_____1966. *Chinese Lineage and Society: Fukien and Kwangtung.* London: Athlone Press.

Goldstone, Jack. 1986. "The demographic revolution in England: A re-examination." *Population Studies* 40: 5-33.

Hajnal, John. 1982. "Two kinds of pre-industrial household formation system." *Population and Development Review* 8: 449-494.

Huang, Philip. 1990. *The Peasant Family and Rural Development in the Yangzi Delta, 1350-1988*. Stanford: Stanford University Press.

Harrell, Stevan. 1987. 'On the Holes in Chinese Genealogies.' *Late Imperial China* 8.2:53-79.

Hsiung Ping-chen. 1994. 'Constructed emotions: The bond between mothers and sons in late imperial China.' *Late Imperial China* 15.1 (June): 87-117.

Jiang Tao. 1993. *Zhongguo jindai renkou shi* (Modern Chinese Population History). Hangzhou: Zhejiang renmin chuban she.

Lang, Olga. 1946. *Chinese Family and Society*. New Haven: Yale University Press.

Laslett, Peter. 1977. "Characteristics of the Western family considered over time." *Journal of Family History* 2.2: 89-116.

_____1983. “Family and household as workgroup and kin group: Areas of traditional Europe Compared.” In Wall, Robin, and Laslett, 513-563.

_____1988. “Family, kinship and collectivity as systems of support in pre-industrial Europe: A consideration of the ‘nuclear hardship’ hypotheses.” *Continuity and Change* 3: 153-175.

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.

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

Levine, David. 1987. *Reproducing Families: The Political Economy of English Population History*. Cambridge: Cambridge University Press.

Liu, Wang Hui-chen. 1959. *The Traditional Chinese Clan Rules*. New York. 1959.

Macfarlane, Alan. 1978. *The Origins of English Individualism: Family, Property, and Social Transition*. Oxford: Oxford University Press.

_____1986. *Marriage and Love in England, Modes of Reproduction 1300-1840.* Oxford: Basil Blackwell.

_____1987. *The Culture of Capitalism*. Oxford: Oxford University Press.

_____1997. *The Savage Wars of Peace: England Japan and the Malthusian Trap.* Oxford: Basil Blackwell.

Plakans, Andrejs. 1984. *Kinship in the Past: An Anthropology of European Family Life, 1500-1900*. Oxford and New York: Basil Blackwell.

Schofield, Roger. 1985. “English marriage patterns revisited.” *Journal of Family History* 10: 2-20.

_____1989. “Family structure, demographic behavior and economic growth.” In Walter and Schofield, 279-304.

Skinner, G. William. 1986. ‘The Population of Sichuan in the Nineteenth Century: Lessons from Disaggregated Data.’ *Late Imperial China*, 7.2:1-79.

Skinner, William G. 1997. “Family systems and demographic processes.” In Kertzer and Fricke, 53-95.

Szonyi, Michael. 2002. *Practicing Kinship: Lineage and Descent in Late Imperial China.* Stanford: Stanford University Press.

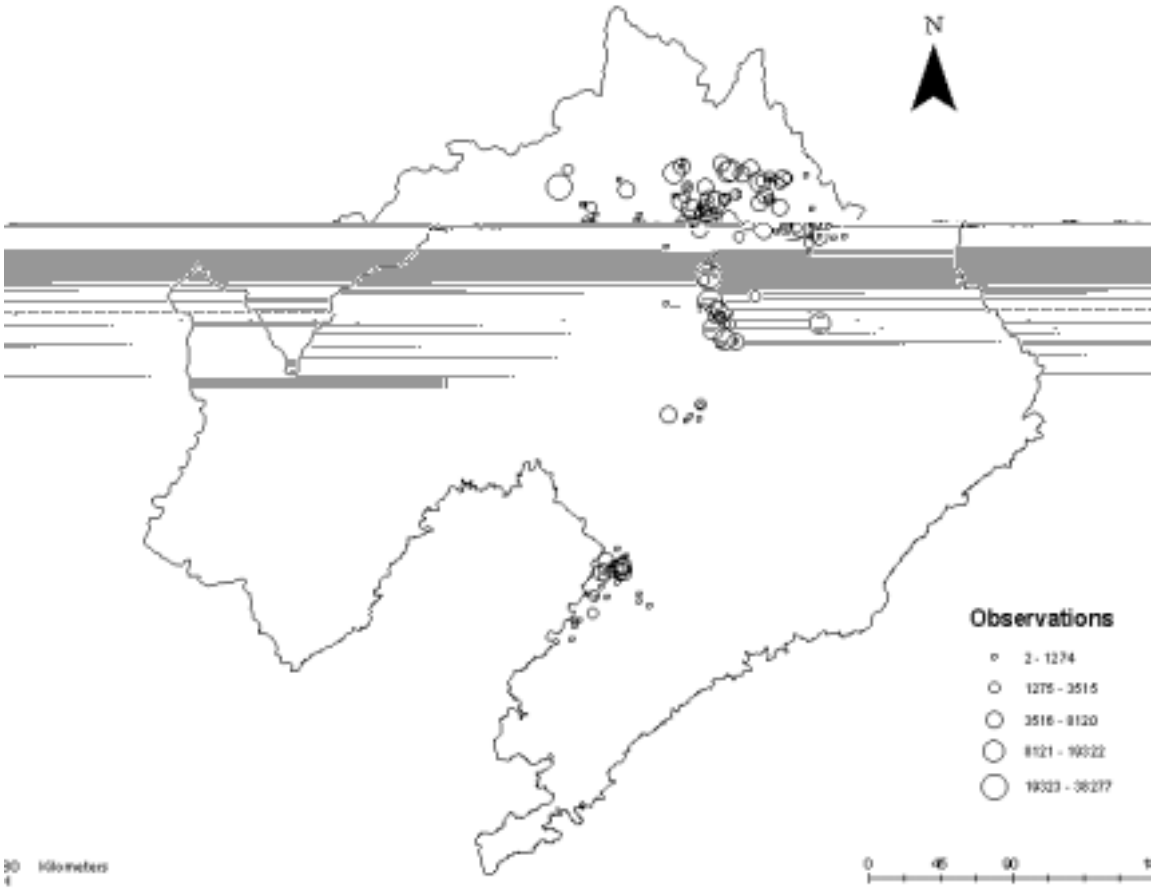
Chia.

Gugone bowuyuan yuankan.

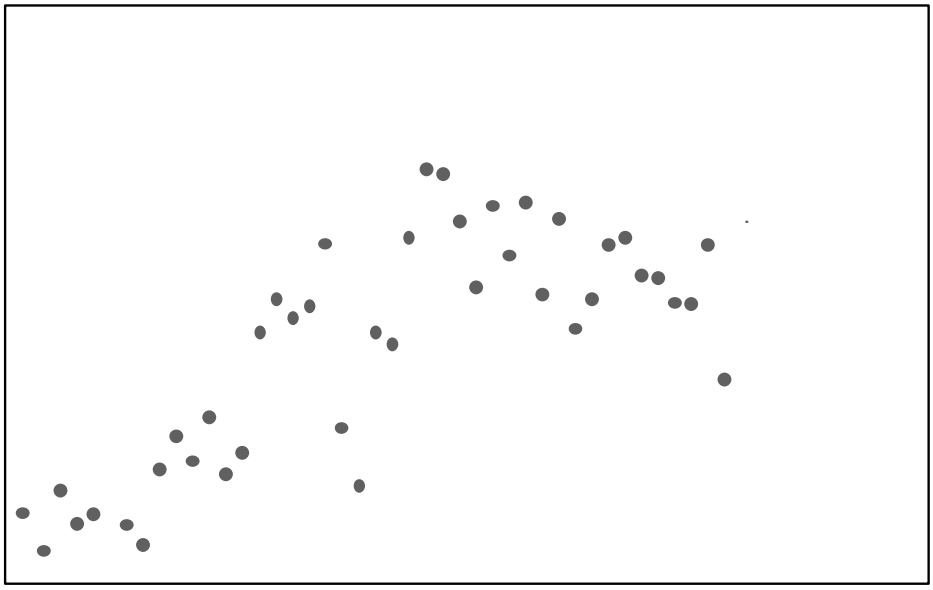
huayan xueshu lunwen heji.

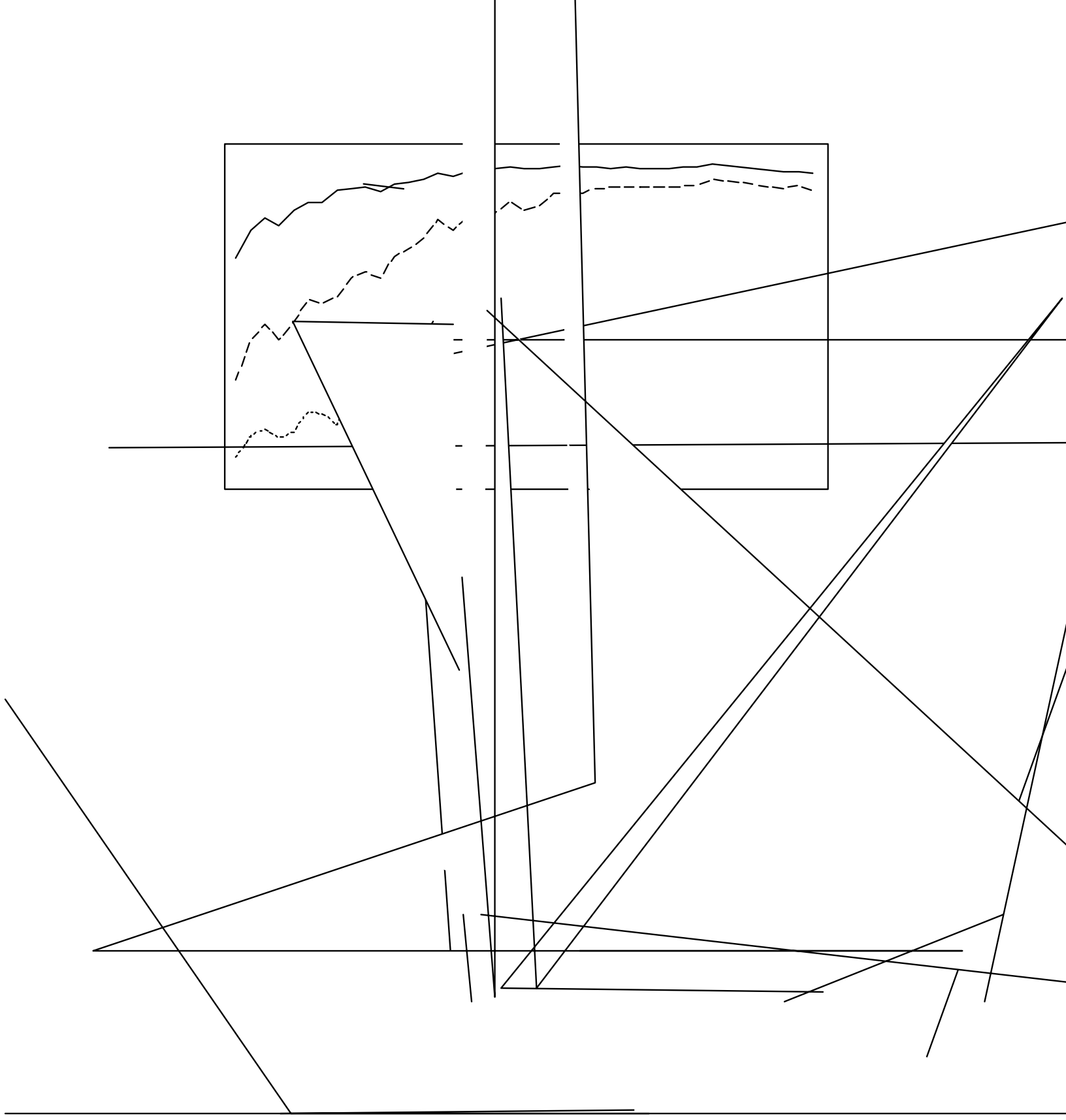
Journriaof Family History.

HistoricriaStudies of ChangicinFertilrsi,.



Map 1 Geographic Distribution of Observations, Liaoning, 1749-1909





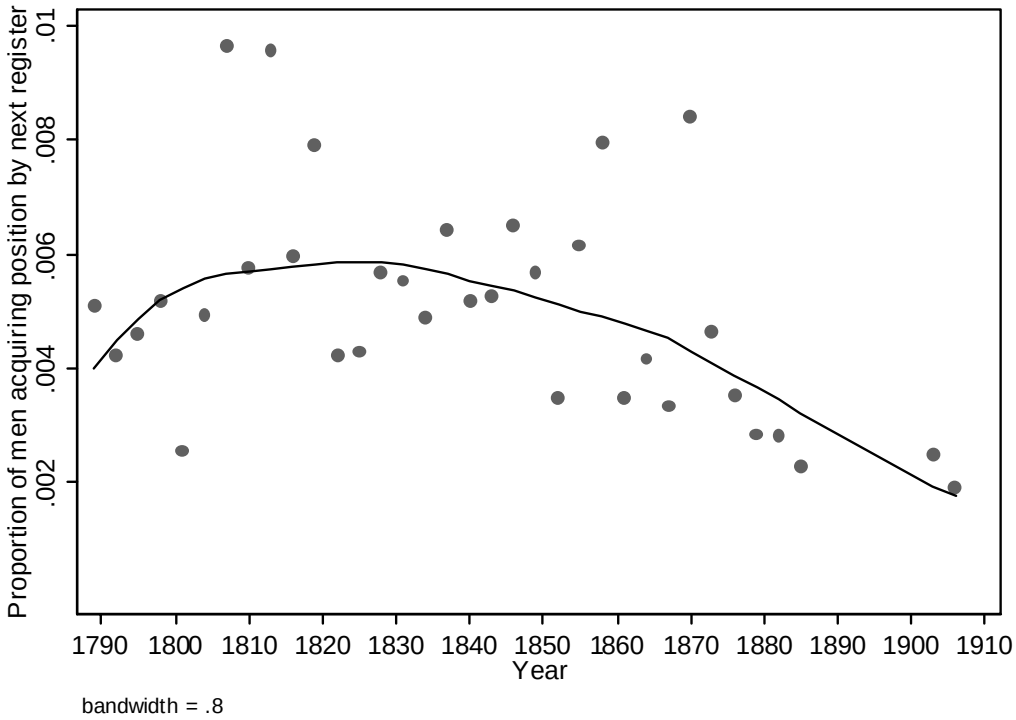
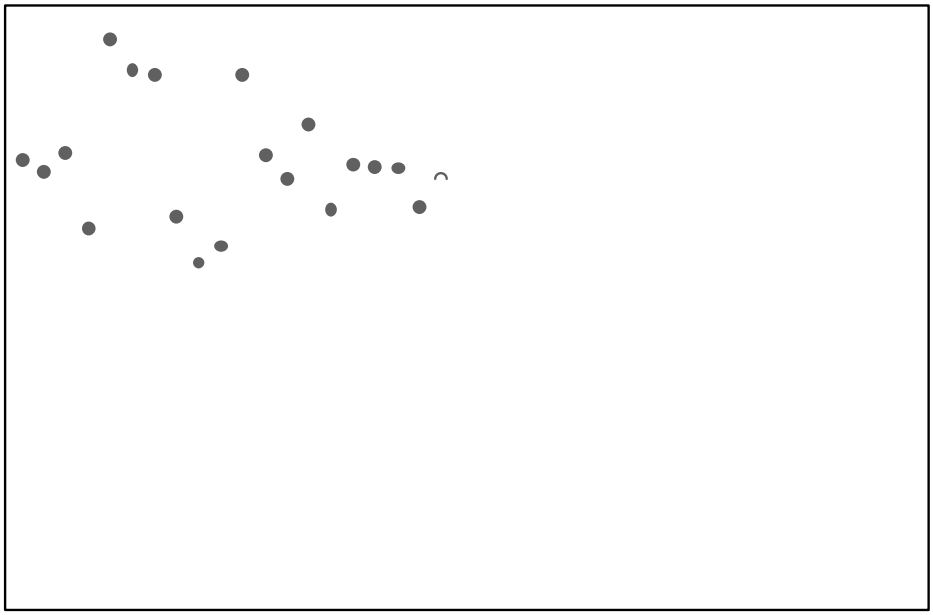


Figure 3 Proportion of adult males acquiring an official position by next register, Liaoning, 1789-1909



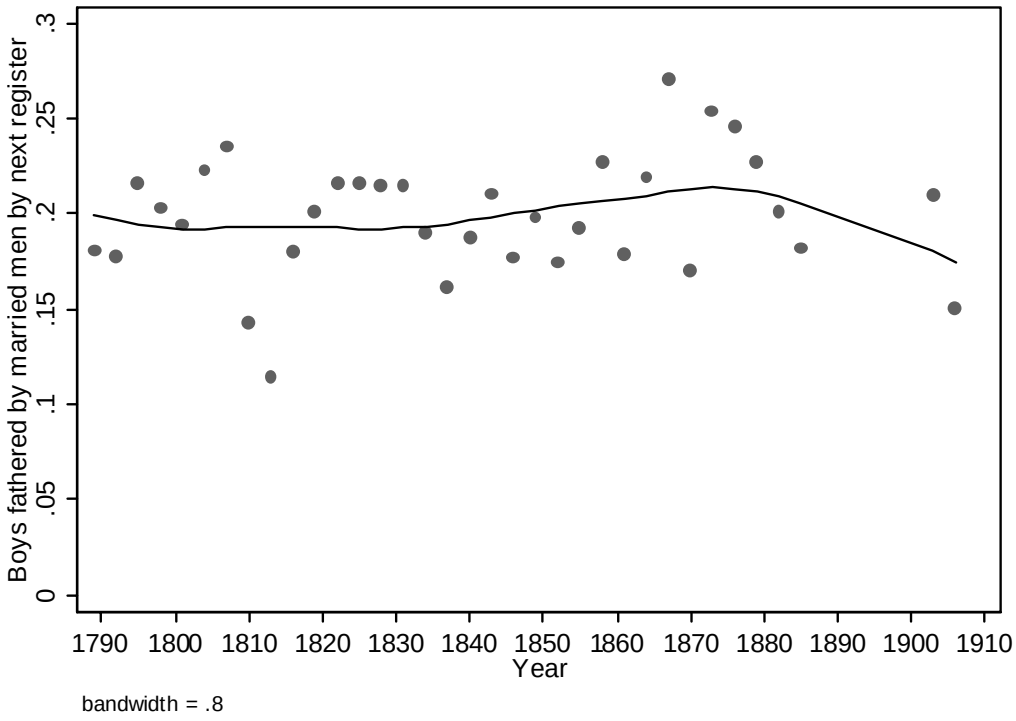


Figure 5 Numbers of boys fathered by married males by next register, Liaoning, 1789-1909

Table 1. Available Household Register Data, by State Farm Population

State Farm Population	Coverage	Household Registers	Observations
Bakeshu	1759-1909	30	40267
Changzhaizi	1768-1909	25	38795
Chengnei	1765-1861	15	29578
Dadianzi	1756-1909	27	64938
Dami	1759-1909	31	25379
Daoyitun	1774-1909	35	118633
Daxintun	1750-1909	27	77694
Diaopitun	1768-1909	26	70153
Feicheng	1756-1909	39	58859
Gaizhou Manhan	1753-1909	20	45043
Gaizhou Mianding	1789-1909	17	22558
Gaizhou	1769-1909	29	42834
Guosantun	1778-1909	32	35073
Langjiabao	1766-1909	25	47340
Nianmadahaizhai	1750-1909	31	52130
Niuzhuang Liuerbao	1777-1906	25	50256
Zhaohuatun	1774-1909	26	50865
Total		534	870,395

Table 2. Means of the variables included in the analysis

Variable	Attainment of position	First marriage	Reproduction
Outcome	0.006	0.221	0.201
Born 1840 or later	0.25	0.35	0.23
Position held by			
Father	0.11	0.11	0.14
Grandfather	0.10	0.11	0.11
Self		0.01	0.05
Brother	0.03	0.02	0.09
Uncle	0.08	0.08	0.11
Non-coresident father's cousin	0.02	0.03	0.03
Coresident father's cousin	0.02	0.02	0.02
Number of kin			
Brothers	1.03	1.01	1.05
Paternal cousins	1.08	1.02	1.15
Coresident paternal second cousins	0.34	0.45	0.33
Non-coresident paternal second cousins	0.67	0.57	0.69
Eldest among			
Brothers	0.66	0.59	0.67
Male paternal cousins	0.41	0.37	0.41
Male paternal second cousins	0.22	0.21	0.22
Males in household	0.28	0.56	0.31
Grandfather not identified	0.16	0.12	0.17
Great-grandfather not identified	0.42	0.34	0.44
Next register 6 years away	0.19	0.19	
Observations	165665	84040	112654

Table 4. Logistic regression of first marriage by next register, never-married Liaoning males, 1789-1909

	Model I	Model II w/ fixed effect of kin group[b]	Model III w/ interactions for birth in or after 1840

Table 5. Poisson regression of number of sons born by next register, married Liaoning males, 1789-1909

	Model I		Model II w/ fixed effect of kin group		Model III w/ interactions for birth in or after 1840	
	Incident Rate Ratio	p-value	Incident Rate Ratio	p-value	Incident Rate Ratio	p-value
Born 1840 or later	1.05	0.01			1.12	0.01
Position held by kin						
Father	1.07	0.06	1.03	0.52	1.07	0.11
Grandfather	1.05	0.03	1.00	0.90	1.07	0.02
Self	1.13	0.01	1.06	0.37	1.16	0.00
Brother	1.01	0.81	0.94	0.28	0.99	0.82
Uncle	1.01	0.76	1.05	0.40	1.00	0.97
Non-coresident father's cousin	1.00	0.96	1.02	0.69	1.02	0.71
Coresident father's cousin	1.04	0.31	1.06	0.24	1.11	0.07
Numbers of kin						
Brothers	1.02	0.01	1.00	0.72	1.03	0.00
Paternal cousins	1.00	0.83	1.00	0.74	1.00	0.59
Coresident paternal second cousins	1.01	0.23	1.00	0.67	1.00	0.69
Non-coresident paternal second cousins	1.00	0.83	1.00	0.99	1.00	0.36
Eldest among						
Brothers	1.02	0.44	1.00	0.91	1.02	0.40
Male paternal cousins	1.03	0.24	1.05	0.17	1.02	0.56
Male paternal second cousins	1.00	0.97	0.97	0.42	1.01	0.76
Males in household	0.92	0.00	0.93	0.01	0.92	0.00
Born 1840 or later *						
Father					0.99	0.94
Grandfather					0.95	0.32
Self					0.83	0.12
Brother					1.06	0.51
Uncle					1.07	0.49
Non-coresident father's cousin					0.96	0.65
Coresident father's cousin					0.89	0.18
Numbers of kin						
Brothers					0.97	0.02
Paternal cousins					0.99	0.56
Coresident paternal second cousins					1.01	0.18
Non-coresident paternal second cousins					0.99	0.24
Eldest among						
Brothers					0.99	0.82
Male paternal cousins					1.05	0.36
Male paternal second cousins					0.96	0.41
Males in household					0.99	0.83
Observations		112654		49566		112654
Log-likelihood		-57929.89		-22079.00		-57917.53
Degrees of freedom		41		23		58

a Dummies for state farm population, five-year age group, next register six years away, grandfather unidentified, and great-grandfather unidentified were also included. To save space, the results are not presented here. See text for details on the definition of each variable.

b The kin group here is defined as consisting of males who have a common paternal great-grandfather.

Last updated 5/23/2014 3:52:02 PM

China Multi-Generational Panel Dataset-Shuangcheng (CMGPD-SC)
User Guide

Version 4

Comments and Feedback

Please send questions, comments, corrections, or suggestions to cmgpd@icpsr.umich.edu.

To receive announcements about new releases of data and documentation, please subscribe to the announcement mailing list at: http://lists.icpsr.umich.edu/mailman/listinfo/cmgpd-announcements.

Recommended citation for User Guide

Wang, Hongbo, Shuang Chen, Hao Dong, Matt Noellert, Cameron Campbell, and James Z Lee. 2013. *China Multi-Generational Panel Dataset, Shuangcheng (CMGPD-SC) 1866-1914. User Guide.* Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

Recommended citation for CMGPD-SC Dataset

Lee, James Z., Shuang Chen, Cameron D. Campbell, and Hongbo Wang. China Multi-Generational Panel Dataset, Shuangcheng (CMGPD-SC), *1866-1914* [Computer file].

Acknowledgment of Support

Please include the fort,owing acknowledgment in any publication, working paper, manuscript or thesis that makes use of the CMGPD-SC data:

Preparation of the CMGPD-SC and accompanying documentation for public release via ICPSR DSDR was supported by the National Institutes of Health, Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) Grant no. R01 HD070985 Multi-generational Demographic and Landholding Data: CMGPD- Contents are solely the responsibility of the authors and do not necessarily represent the official views of the NICHD.

If you make use of the User Guide or Dataset in a publication, working paper, thesis or dissertation, please email the complete reference to cmgpd@icpsr.umich.edu so that we can include it in our reports on usage. We would be grateful as wert, for electronic copies of the final publication.

Table of Contents

DRAFT

List of Figures

Last updated 5/23/2014 3:52:02 PM

Introduction

The China Multi-Generational Panel Dataset - Shuangcheng (CMGPD-SC) provides longitudinal individual, household, and community information on the demographic and socioeconomic characteristics of a resettled population living in Shuangcheng, a county in present-day Heilongjiang Province of Northeastern China, for the period from 1866 to 1926. The dataset includes some 1.3 million annual observations of over 100,000 unique individuals descended from families under the Eight Banners system who were relocated to Shuangcheng in the early 19th century. Table 1 categorizes these individuals according to their original place of origin as metropolitan bannermen or *jingqi* ( ) from Beijing and Rehe (present-day Chengde); rural bannermen or *tunding* ( ) from the provinces of Liaoning and Jilin; and floating bannermen or *fuding* ( ) largely from the province of Liaoning. These three categories accounted for the majority of the registered residents in Shuangcheng during this period with different socioeconomic statuses and entitlements to land (Chen 2009, 106-138).

Table 1 Number of Registers, Observations, and Individuals by Banner Population

		Qing						1926
Dataset		Start	End	Interval b/w registers	No. of Registers	Obs.	Individuals	Obs.
Metropolitan Bannermen								
101	Plain Yellow	1866	1912	1 year	40	70,897	4,838	
102	Bordered Yellow	1866	1912	1 year	41	66,853	4,391	
Subtotal:					*81*	*137,750*	*9,229*	
Rural bannermen								
103	Plain White	1868	1910	1 year	26	121,401	11,006	
104	Bordered White	1866	1911	1 year	36	217,145	14,054	9,916 [1]
105	Plain Red	1866	1913	1 year	32	165,535	12,320	4,184
106	Bordered Red	1866	1912	1 year	31	195,853	15,390	
107	Plain Blue	1866	1911	1 year	30	185,056	14,547	6,675
108	Bordered Blue	1866	1909	1 year	34	195,587	12,804	
Subtotal:					*189*	*1080,577*	*80,121*	20,775
Floating Bannermen								
111	Plain White	1867	1909	1 year	12	17,253	2,104	
112	Bordered White	1870	1909	1 year	11	38,418	5,175	
113	Plain Red	1867	1909	1 year	9	6,930	1,962	
114	Bordered Red	1867	1909	1 year	12	15,283	2,379	
115	Plain Blue	1867	1909	1 year	13	35,832	4,773	
116	Bordered Blue	1867	1909	1 year	11	14,783	2,147	
Subtotal:					*68*	*128,499*	*18,540*	
Total:					338	1,346,826	107,890*	20,775

Source: CMGPD-SC, 1866-1913, 1926.
Note: [1] Although the actual year when this Bordered White register was compiled is unknown, it is assumed to be compiled in 1926.
*The total number of individuals here is the sum of the number of individuals in the three population categories: metropolitan, rural, and floating bannermen. Because some individuals changed their population category in the period covered by the household registers, they were counted twice in table 1. The number of unique individuals is 107,551.

The CMGPD-SC, like its Liaoning counterpart, the CMGPD-LN,[1] is a valuable data source for studying longitudinal as well as multi-generational social processes. In addition, the CMGPD-SC has three further valuable qualities which make it a superior resource for social science and health research. Fi[( )theTJETBT1 0 0 1 354.68 335.66 Tm[(so)eTJETBT

and descriptive statistics. Part four summarizes the spatial, demographic and socioeconomic characteristics for each of the population categories.

1 Shuangcheng Banner Population and Land Registers

1.A The Shuangcheng Settlement

The population records in the CMGPD-SC are transcribed from the Eight Banner population registers preserved in the Liaoning Provincial Archives, which are

Map 1 Contemporary Shuangcheng County with CMGPD-SC Villages[3]

Two hundred years ago, however, Shuangcheng was a largely uninhabited grassland plain. Settlement and transformation only began in 1815, when the Qing government in order to relieve the fiscal hardship of supporting the banner population in Beijing, initiated 0 1 ETBT(h)-343.95 -3(5.9 476 598.T)4(ve3252.2c43.14.29 Tm[(e)4(d)] TJ15 26BT

In 1819-1820, the Qing court relocated 2,000 more banner households from Liaoning and Jilin to Shuangcheng. It therefore expanded the Shuangcheng state farm and built 80 more villages, 40 to the west and 40 to the east of the original settlement. Accordingly, the original 40 villages were named central *tun* and the two new settlements were named right and left *tun* respectively. As Map 1 and Map 2 show, compared to the central *tun*, the distributions of the villages in the right and left *tun* are less regular. This is because the hills and marshes on the peripheries of the plain prevented the government from laying the villages out as symmetrically as originally planned. The 2,000 households of rural bannermen from Liaoning and Jilin settled in the right and left *tun*.

Then, beginning in 1824, the government started to relocate metropolitan bannermen to Shuangcheng. From 1824 to 1838, a total of 698 households of metropolitan bannermen moved from Beijing and Rehe and settled in the 40 villages of the central *tun*. As such, the Shuangcheng state farm accommodated a total of 3,698 households of official immigrants.

The Shuangcheng settlement created a heterogeneous population in terms of place of origin, life style, and ethnicity. The immigrants came from 19 different places. While the left and right *tun* residents were exclusively rural bannermen, the residents of the 40 villages of the central *tun* were a mixture of urban migrants and rural settlers. Moreover, since the government settled immigrants from the same place of origin into different villages, the population in each banner village was also extremely heterogeneous (Chen 2009, 89-93).

Compared to that of the CMGPD-LN population, the ethnic composition of the CMGPD-SC is also far more complex. The population consisted of a total of six ethnic groups: Manchu, Mongol, Han, Xibe, Baerhu, and Taimanzi.[6] Therefore, the Shuangcheng banner villages were also ethnically heterogeneous. If we consider the metropolitan and rural bannermen only, the residents of 4 of the 120 villages belonged to as many as five ethnic groups. Eighty-seven villages had three or four ethnic groups. Only eleven villages consisted of a single ethnic group (Chen 2009, 96).The state divided these official immigrants and other unofficial immigrants to the Shuangcheng area into four population categories: metropolitan bannermen and rural bannermen who were official immigrants; floating bannermen, who shared the places of origin of rural bannermen but moved to Shuangcheng without official order; and civilian commoners who were unofficial immigrants. Moreover, by assigning these population categories different entitlements to land, the state constructed a social hierarchy whereby the metropolitan and rural bannermen became the local haves and the floating bannermen and civilian commoners became the local have-nots (Chen 2009, 106-138).[7] Among the haves, metropolitan bannermen enjoyed the greatest benefits, with land allocations twice that of rural bannermen, better housing, and complete assistance in farming.

The state defined population categories in Shuangcheng, while mainly

[6] While Manchu, Mongol, Han, and Xibe were common ethnic categories in the Eight Banners. Baerhu was a Mongol ethnic group originated in today's Mongolia.

[7] Notwithstanding the ban on Han immigration to the region, Shuangcheng had a growing civilian commoner (*minren*) population.

reflecting their place of origin and official immigration status, also correspond to the ethnic hierarchy in the Eight Banners. The metropolitan bannermen virtually consisted of only Manchu and Mongol--the top two ethnicities--with an overwhelming proportion (82.6%) belonging to the Manchu ethnic group. Only a handful of households of metropolitan bannermen belonged to the Xibe ethnic group. The rural bannermen consisted of all six ethnic groups, with only 43.5% of the population belonging to Manchu. The floating bannermen had ethnic composition similar to the rural bannermen, but only 34.1% of the population belonged to Manchu.

From the time of settlement to 1931, the banner population in Shuangcheng was administered by a banner government. At first, this banner government was in charge of all bannermen affairs including population registration, land allocation, taxation, public security, and such civil affairs as land disputes. In 1882, however, the government established a parallel civilian administration to accommodate the increasing civilian population in the area. In 1909 the government further upgraded this county-level civilian government to a prefecture-level one and established a banner sub-office within the civilian government to administer the banner population (Chen 2009, 161-175, Ren 2012).

From 1820 to 1869, the Shuangcheng banner administration had three sets of eight banners with distinct offices located in the central, right, and left *tun*, respectively. In 1869, the state reorganized the banner administration, consolidating the three sets of eight banners into one. Consequently, a banner consisted of 20 villages. As Map 1 and Map 2 show, the Plain Yellow banner and Plain Red banners administered the immigrants in the 20 villages west of the seat of Shuangcheng, with the Plain Yellow banner administering the metropolitan banner households and the Plain Red banner administering the rural banner households. Similarly, the Bordered Yellow and Plain white banners administered the immigrants in the 20 villages east of the seat of Shuangcheng, with the Bordered Yellow banner administering the metropolitan and the Plain White banner administering the rural banner households. The Bordered Red and Bordered Blue banners administered the immigrants living in the right *tun*, and the Plain Blue and Bordered White banners administered those living in the left *tun*.

The banner population in the CMGPD-SC dataset differs from that in the CMGPD-LN with respect to administration, origin, and socioeconomic status. Although both populations were banner populations, the CMGPD-SC population belonged to the *baqi dutong* (. banner command system and was therefore classified as regular bannermen whose main responsibility to the state was to provide military service. The CMGPD-LN population, by contrast, belonged to the Imperial Household Department (*neiwufu*) and was responsible for providing the Qing court with agricultural produce and a wide variety of specialized labor including military service, but also much else. As a result the CMGPD-SC bannermen, including those who originated from Liaoning, belonged to a totally different administrative system.

Compared to the CMGPD-LN population, the CMGPD-SC population had higher socioeconomic status. Not only were metropolitan bannermen from Beijing considered an elite group even among Eight Banner populations, the rural bannermen from Liaoning in the CMGPD-SC enjoyed greater economic benefits than the

CMGPD-LN population. In Shuangcheng, both metropolitan and rural bannermen received state-allocated land without paying tax or rent. They could use this land as their own property and pass it down to their descendants. These economic benefits explain significant differences in the registration behavior between the CMGPD-SC and CMGPD-LN populations. As we will discuss in detail later, the CMGPD-SC population, for example, tended to register their children much earlier than the CMGPD-LN population, thus making the CMGPD-SC dataset more suitable for the study of fertility. In the same light, differences in socioeconomic status between the CMGPD-LN and CMGPD-SC populations also explained the quality of recording of disability status. Because the CMGPD-SC population did not provide corvee labor to the state, the recording of their disability status was sporadic.[8]

1.B Population Registration in Shuangcheng

The banner population registers in Shuangcheng were compiled and maintained by the local banner government. At the time of relocation of metropolitan and rural bannermen, the state transferred their registration records from their places of origin to Shuangcheng. The Shuangcheng banner government then compiled these r organized by their new residence and banner affiliation, preserving all information transferred from their place of origin (See Figure 2). Entries in each register were grouped first by village, then by household group (*yihu*) and then by household. Each entry first recorded information about the household head: his place of origin, ethnicity, original banner affiliation, occupation, name, age, and any vital demographic event that had happened since the last update. The registers subsequently recorded the relationship, name, age, and occupation of the immediate family members (wife, children, and parents) and then of any other relatives living with him. The register indexed all household members by their relationship to the principal adult male.

Throughout the history of the Shuangcheng state farm, the banner population registers served as the official references for administration and land allocation. The state inscribed metropolitan and rural bannermen into different registers, documenting their different membership and entitlement rights to state land. Only those males recorded in the metropolitan and rural registers were eligible for state land allocation. Because population registration played an important role in land allocation, the Shuangcheng local government updated registers annually. In the eleventh month of each year, the local government would compile a clean copy of the updated registers and send it to the provincial government for review. As a result, compared to banner population registers in other places, which were updated triennially,[9] the detail and completeness of recording in the Shuangcheng banner population registers were of higher quality.

[8] Due to this reason, the CMGPD-SC data do not have the variable DISABLED, which was included in the CMGPD-LN data.

[9] During the Qing, the state regulation stipulated that banner population registers be updated every three years. The banner population registers in Shuangcheng are the only annually updated registers that we have found.

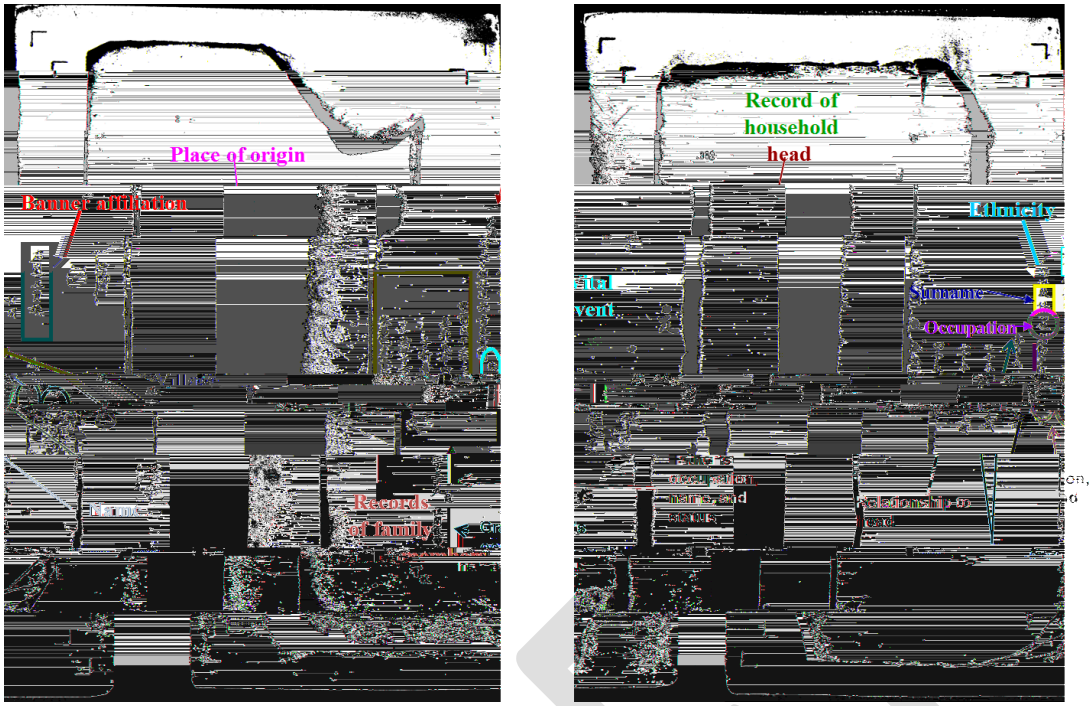


Figure 2 Sample Pages of CMGPD-SC Population Register

1.C Land Allocation in Shuangcheng

Upon settlement in Shuangcheng, the government allocated land to metropolitan and rural banner households, thereby creating a state-mandated social hierarchy. To do so, the government first made the household the unit of land allocation and appointed the registered household head to be in charge. Then the government allocated one plot of land to each household (Figure 3 and Figure 4). The 64.4 hectare plots allocated to metropolitan banner householdswere nearly twice the size of those allocated to rural banner households, which were 34 hectares. Moreover, within each population category, the government tried to maintain an equal distribution of land, stipulating that one household could only own one plot of allocated land (Chen 2009).

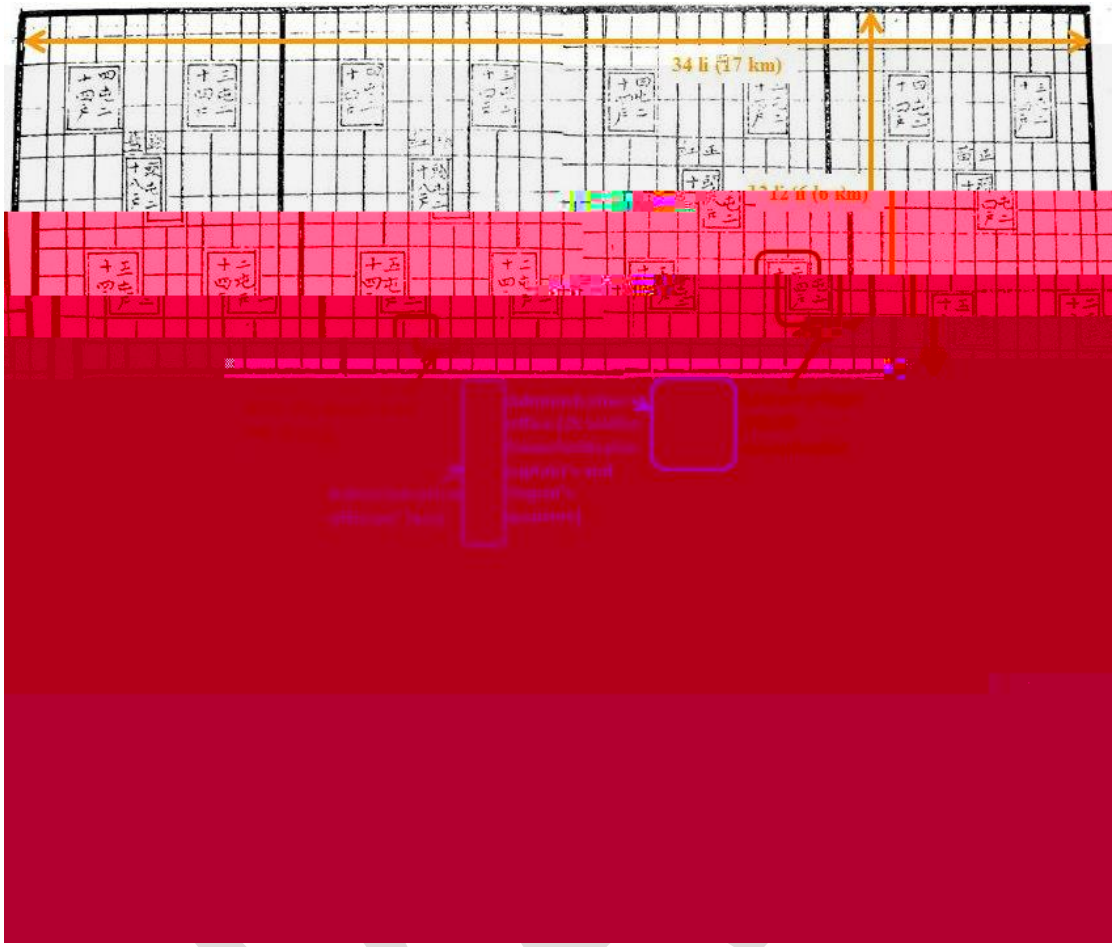


Figure 3 Banner Village Blueprint with Gridded Land Plots (ca. 1820)[10]

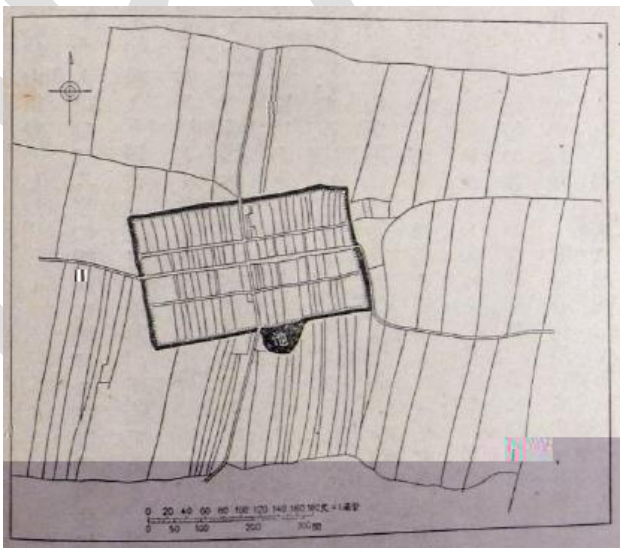


Figure 4 Land plot map for a single banner village, ca. 1940[11]

In addition to their allotment of state land, Shuangcheng residents also cleared other land privately. Beginning in 1844, the government gradually inventoried privately-cleared land and registered these land holdings. The residents then paid rent

[10] SCPTTJL 1990.
[11] (Komekura 1941, 142).

linking land ownership records in 1870, 1876, 1882, 1887, and 1889 to population records from the same or adjacent years.[13] For each banner, we first used a computer program to link the records of land ownership to the population dataset by merging the year of registration, village of residence, and name of the land owner. We then hand-linked the remaining observations to the population dataset, adding land holding data to the socio-demographic data in the CMGPD-SC for these 13,155 individuals.

2 CMGPD-SC Register Data

2.A Overview

Like its Liaoning counterpart CMGPD-LN, the CMGPD-SC is also suited to the study of a wide variety of topics in demography, family dynamics, and social stratification. It provides rich data on a largely closed population, which are transcribed from mainly annual administrative registers. These uniformly structured registers allow us to link individuals and their families over time, effectively producing the longitudinal component of the CMGPD-SC dataset. As a result, the CMGPD-SC follows individuals prospectively with time-varying characteristics available annually for metropolitan and rural bannermen, who comprise over four-fifths of the CMGPD-SC population, and triennially for f bannermen, who comprise almost all the remaining population. The timing of key economic, social, family, and demographic events and transitions can thus be ascertained. Contextual information as to community and household are also available at regular intervals. As previously highlighted, the CMGPD-SC also contains rare longitudinal information on property in the form of individual landholding, which distinguishes it from many other comparable historical data sources, including the CMGPD-LN.

Alongside its Liaoning counterpart, women were recorded in detail when they were wives or widows. The CMGPD-SC records daughters more completely than the CMGPD-LN or any other demographic source for a non-elite pre-twentieth century Chinese population. This is particularly true for the metropolitan bannermen.

With such features, the CMGPD-SC stands as a rich source for the study of kinship networks and multi-generational processes, a signature strength shared by the CMGPD-LN (Mare and Song 2012, Song, Campbell and Lee 2012). Specifically, through manual and automated linkage of individual records and clusters of records in the original registers, the dataset tracks individuals and families across multiple generations, and reconstructs networks of paternal kin living outside the household. Thus, the CMGPD-SC supports analysis of associations between more distant relatives and within much more broadly defined kin groups, well beyond the more canonical analyses of associations of characteristics, transitions, and outcomes across the life course, between parents and children, or between siblings. Given the kin-centered nature of much East Asian society, this feature is of special relevance for the study of East Asian social and economic behavior.

[13] The land records are linked to the adjacent years if the population register of the same year of the land register is missing.

The most notable limitation of the CMGPD-SC is the relatively short time span it currently covers 60 years compared to 160 years for the CMGPD-LN. As a result, its strength for studying multi-generational processes is more limited than the CMGPD-LN. Other limitations the CMGPD-SC also shares with its Liaoning counterpart include: the omission of boys who died in infancy or early childhood as well as the absence of recording for many daughters, albeit to a lesser degree than the CMGPD-LN. In addition, as in the CMGPD-LN, socioeconomic status measures are available only for males who hold official positions or titles.

Below, we highlight the major advantages of the CMGPD-SC and discuss in detail these strengths and limitations and provide examples of applications to the study of key areas in demography, family dynamics, and social stratification.[14]

2.B Strengths

2.B.I Prospective

Similar to the CMGPD-LN, the data in the CMGPD-SC are prospective and superior to the retrospective data recorded in lineage genealogies to the extent that selectivity bias and recall errors are minimal. Each register describes conditions around the time of its compilation and records for each individual any exits due to death, out-marriage, emigration, and illegal departures that occurred during the last year for metropolitan and rural bannermen or last three years for floating bannermen. The registers also document detailed relationship to the household head for each individual. For adult males, the registers further record official administrative statuses.

2.B.II Longitudinal

The CMGPD-SC follows individuals through the life course at annual or triennial intervals, which in turn allows us to determine for most individuals the timing of their entrance into the dataset through marriage or birth by comparison between registers. Timing of the attainment of official position and retirement, can also be similarly inferred by comparison of statuses in adjoining registers. Note that the original registers, which are akin in format and organization to annual censuses, are not longitudinal by themselves. However, the fact that the Shuangcheng registers list individuals in roughly the same order in successive registers allows for easy record linkage between adjacent registers. Life histories of the CMGPD-SC individuals can be analyzed using such techniques as survival analysis. The longitudinal data in the CMGPD-SC also allows for evaluation of new techniques for age-period-cohort analysis.

We use computer programs to aggregate the links between pairs of records in adjacent registers and create unique identifiers to group the records in different registers that correspond to an individual. Automated linkage of records of the same individual in different registers is the basis of the variable PERSON_ID, which identifies all the observations of a person across different registers. PERSON_ID is also used as the basis to create links between individuals through such variables as

[14] Users are advised to refer to - -LN User Guide (Lee, Campbell and Chen 2010).

WIFE_ID, HUSBAND_ID, FATHER_ID and so forth. Figure 6 shows the population-specific proportions of observations successively linked to the previous register across years, which are typically well over 90 percent.

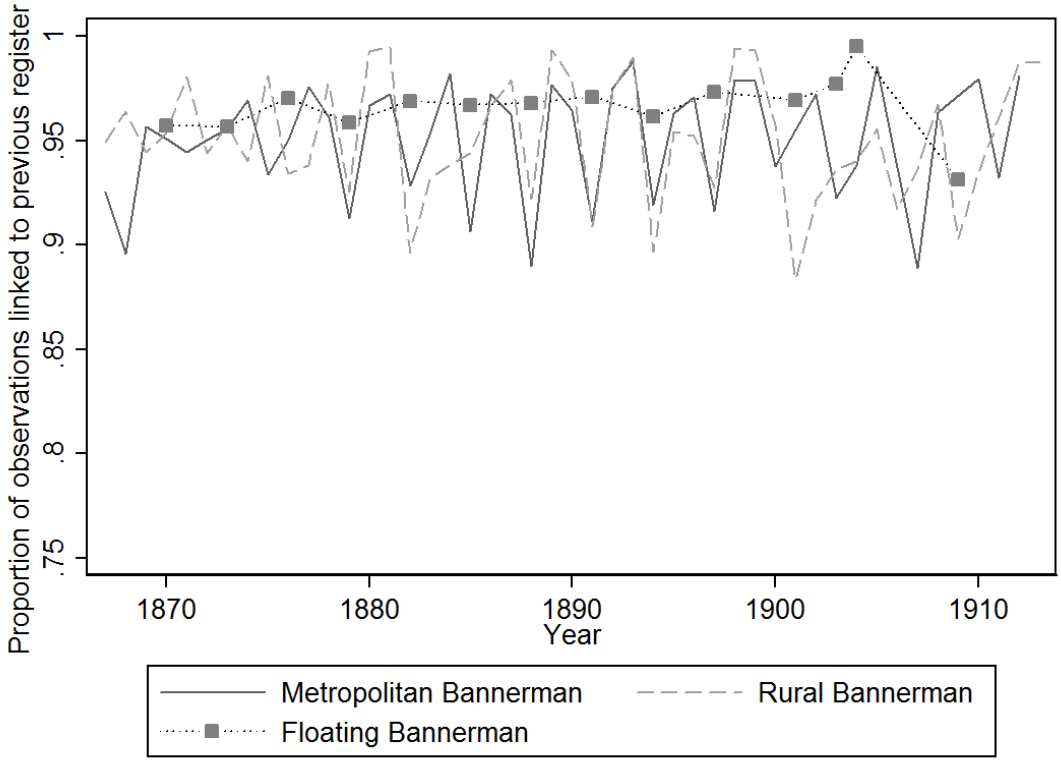


Figure 6 Proportion of observations linked to previous register.

2.B.III Multi-generational

A distinguishing strength of the CMGPD-SC lies in its potential for examining multi-generational processes and networks of distant kin. Children have been linked to their parents through automated record linkage, and those links aggregated to reconstruct descent lines and kinship networks. The basic procedure has been to chain together links between fathers and sons from the raw data to identify grandfathers, great-grandfathers, and earlier male ancestors. This intergenerational linkage is the basis of such linking variables as FATHER_ID, MOTHER_ID, and GRANDFATHER_ID, kin count variables such as UNCLE_COUNT, AUNT_COUNT, BROTHER_COUNT, SISTER_COUNT, MALE_COUSIN_COUNT, FEMALE_COUSIN_COUNT, and a variety of kin or descent group identifiers in the dataset.

Overall, we were able to link 76.8 percent of all CMGPD-SC males to their fathers;[15] 63.7 percent of these 45,688 males to their grandfathers; and 30.3 percent of these 29,090 males to their great-grandfathers. In each case, floating bannermen were the hardest to link with proportions generally half that of metropolitan and rural bannermen. Figure 7 summarizes the proportions of male children in successive decades for whom specified paternal ancestors have been identified. The proportion of children with an identified great-grandfather increases in later registers and reaches 55.3

[15] The corresponding proportion of males who can be linked to their father in the CMGPD-LN is 87.81.

percent by the beginning of twentieth century. Indeed, for children born after 1900, we can link as many as 9.3 percent all the way back to their great-great-grandfather.

Figure 7 Proportions of Children for Whom Specified Paternal Ancestor Can Be Located in the Registers

These constructed paternal pedigrees allow for measurement of networks of paternal kin. Theoretically, relationships between family members specified in the earliest registers allow for extension of pedigrees by inferring common descent from ancestors whose death preceded the earliest available register. For example, adult men identified as cousins in the earliest register must have had a common paternal grandfather, their sons share a common great-grandfather, and their grandsons share a common great-great-grandfather. Adult men listed as second cousins in the earliest register must have a common great-grandfather, and so on. Through such linkage, it will be possible to divide the population into groups defined according to common descent from a founder, allowing for the study of intergenerational transmission and intragenerational correlations of specific outcomes or characteristics.

Despite its relatively shorter time depth compared to CMGPD-LN, CMGPD-SC is one of only a few datasets that trace a resettled community due to policy concern for half a century or more (Clark, Colson, Lee, and Scudder 1995) The

and world histories and has been an increasingly common phenomenon in the last half-century in both developing and industrialized countries (Colson 1971, Lee 1978, Cernea and Guggenheim 1993, M. Cernea 1985, Oliver-Smith 1982). However, in many cases, the efforts to resettle these populations failed, as the immigrant societies failed to consolidate. Even for the few successful cases, due to the scarcity of detailed individual level or household level records, the situation of their long-term

term health consequences of resettlement and migration for children with CMGPD

preliminarily studied the long

LN (Dong and Lee 2012), its lack of information on wealth, such as land holding, largely limits the possibility of studying the long-term socioeconomic development after resettlement and migration. The CMGPD-SC, however, not only preserves relevant information for the first generation immigrants but also for the second and third generations.

Moreover, since the CMGPD-SC starts with an initial state of limited inequality in a frontier population, which was deliberately categorized by the state into three major groups differing in social origin and property, the possibility of following the evolution of socioeconomic inequality across generations from a well-defined origin will facilitate research beyond standard intergenerational mobility.

2.B.IV Closure

Similar to the CMGPD-LN, the population covered by the CMGPD-SC is largely closed in the sense that for males and married or widowed females, there was relatively little out-migration. Exits were annotated in the original registers, providing the basis for construction of flag variables such as NEXT_ABSCONDED, NEXT_DIE, NEXT_MARRY, and NEXT_REMARRY that indicate an exit between the current register and the next available register.

Changes in location within Shuangcheng can be detected by comparing UNIQUE_VILLAGE_ID for the same individual across adjacent registers. In the CMGPD-SC, there are 9,429 cases of residential mobility within the boundary of Shuangcheng, involving about 3,000 individuals.

Closure to out-migration means that not only can most males and married or widowed females be followed across their life course until they die or the registers end, but also that families can be followed across generations. When individuals left Shuangcheng entirely, their departure was supposed to be annotated in the register, allowing their observations to be censored. NEXT_ABSCONDED indicates whether an individual was annotated in the next register in the CMGPD-SC as absconded. Such cases of annotated abscondance in the CMGPD-SC, however, were rare compared to the CMGPD-LN. Only 122 individuals in Shuangcheng were annotated *tao*). Yet, about two thousand individuals appear to have disappeared in the sense that they were present in one annual or triennial register but missing from the next *adjacent* one in the dataset, without any annotation of exit, such as death or out-marriage.[16]

2.B.V Land Register

The most innovative feature of the CMGPD-SC relative to the CMGPD-LN and indeed to most other publicly available historical population databases is that it provides longitudinal data on wealth in the form of records of individual landholding. Amount, category, and type of land are recorded at six points in time: 1870, 1876, 1882, 1887, 1889 and 1906.

[16] These cases are essentially different from those whose records ended without any annotation of exit due to the fact that the register recorded their exit is missing.

These landholding data allow for measurement of wealth differences in demographic behavior and other socioeconomic attainment. More importantly, the measurement of landholding at multiple points in time allows for the study of wealth stratification by examining the community, household, and individual characteristics associated with subsequent increases or decreases in landholding. While some important contemporary longitudinal studies offer wealth measures at multiple points in time, only the Panel Study of Income Dynamics and the Wisconsin Longitudinal Study come close to the CMGPD-SC in terms of generational depth and detail on kinship. Arguably, for anyone with a general interest in demographic and wealth stratification processes, the CMGPD-SC is an important complement to these contemporary sources by virtue of its focus on a preindustrial, non-Western population. The detail on wealth makes the CMGPD-SC ideal as a source for multi-generational studies of inequality of the sort called for by Robert Mare (2011).

2.B.VI Multilevel

The CMGPD-SC data are also hierarchical in the sense that individuals in the CMGPD-SC are embedded in multiple concentric or in some cases crosscutting layers of context: residential household, household group, paternal descent group, community, village, and banner. In some cases, contexts overlap. In comparison with the case of Liaoning, descent groups within the villages covered by the Shuangcheng registers tended to be more diverse. Another unique characteristic of the Shuangcheng settlement is that certain pairs of banner populations always co-resided in the same villages (See UNIQUE_VILLAGE_ID below for details). Not only do the data allow for examination of how measured characteristics of these different layers of social organization affect individual demographic and social outcomes, the data also allow for application of hierarchical models and other advanced techniques to measure or account for otherwise unmeasured variation at different levels.

For example, HOUSEHOLD_ID identifies observations associated with a particular household in a given year, UNIQUE_VILLAGE_ID identifies all of the observations associated with a particular village, and DATASET identifies the observations associated with a unique group defined by both banner affiliation and place of origin (e.g. Plain Yellow. See Table 1 for a list of all such groups). As previously pointed out, village and banner population can overlap in the CMGPD-SC. Kin group identifiers in the kinship file also allow for grouping of observations by paternal descent group. However, since the CMGPD-SC only covers a single county, it does not have the equivalents of REGION and DISTRICT in the CMGPD-LN.

2.C Limitations

Lee, Campbell, and Chen (2010) have identified four major limitations of the register data in the CMGPD-LN, namely omission of children, missing registers, biased timing and occurrence of events, and coverage and representativeness (29-31). The CMGPD-SC also suffers from these limitations, albeit to a lesser degree. Also, the quality of the registration of vital events also varies among the three population categories in CMGPD-SC; while that metropolitan and rural bannermen is high, that of the floating bannermen is low. Due to this fact, users should exclude floating bannermen from analyses of demographic behavior.

2.C.I Omission of Children

Incomplete fertility histories for individuals due to the underreporting of infants, young boys, and daughters in the original registers are a serious problem for the CMGPD-LN. The CMGPD-SC is also affected by this problem. In contrast to European vital or population registers, both Liaoning and Shuangcheng registers record persons, not *births*. If a newborn child died before the next annual update, he or she would not appear in any register.

There are differences in reporting completeness for girls and boys, and acorss the three soci-economic groups. The most privileged group in the CMGPD-SC, the metropolitan bannermen, show little evidence of underreporting male infants and boys, possibly because of closer surveillance over this group by the banner administration combined with annual population registration. There is no clear evidence that metropolitan bannermen parents did not register their boys until they had survived to around age 5 or later as in the CMGPD-LN. If we limit the calculation to those who first appeared in a population register after 1866, 8.84% of metropolitan banner males were registered for the first time when they were infants (i.e., at 1 *sui*) while most of them were registered for the first time at 2 *sui* (35.92%) or 3 *sui* (23.7%). However, the percentages for metropolitan banner females are half that of males, a sign that underreporting is a more serious problem for girls. In contrast, the corresponding percentages for rural banner males are much smaller (1.0% at 1 *sui*, 11.2% at 2 *sui*, and 19.0 at 3 *sui*). About one fifth of them were registered for the first time at 5 *sui*.[17] The underreporting of daughters is also more of a problem for rural bannermen. Finally, because their status as unofficial immigrants and their lack of entitlement rights of allocated land, the underreporting of infants, young boys, and daughters is most severe for floating bannermen.

Fertility estimates based on births inferred from the records of children are thus incomplete, and require adjustment based on assumptions about the sex ratio of live births and the level of male mortality in infancy and early childhood. Using the CMGPD-LN, Lee and Campbell (1997, 65-70) estimate that -third of all female deaths [in the first year of life] in rural Liaoning and even 3 to 4 percent of likely to have been the direct result of unrecorded deliberate discrimination. Interpretation of results from event-history analysis of fertility should account for the possibility that apparent fertility differentials may also therefore reflect differentials in infant or child mortality, especially by sex.

2.C.II Missing Registers

Not all annual or triennial Shuangcheng registers were available when the CMPGPD-SC was created. Some registers were missing because they were destroyed by fire, bookworms, or are otherwise unavailable. As indicated in Table 1, the coverage of population registers is more complete for metropolitan bannermen than for rural bannermen and for rural bannermen than floating bannermen.

Each missing register in Table 1 means that for that population we have no data in the CMGPD-SC, for the missing year if they are metropolitan or rural

[17] Missing rural banner registers may also drive this pattern. As already mentioned, the recording of metropolitan bannermen was significantly more complete.

bannermen, or for three years if they are floating bannermen. Individuals whose death, out-marriage, or other exits were recorded during these periods will therefore disappear without explanation from the CMGPD-SC. For example, if all the annual registers from 1870 and 1880 survive for a particular metropolitan or rural population with the exception of 1879, for a widow who was alive from1870 to 1878, annotated as dead in the missing 1879 register, and therefore not included in 1880 there is no way of determining whether she died, remarried out, or otherwise emigrated from the household. In the CMGPD-SC there are about 12,500 such right-censored individuals due to missing registers: 427 metropolitan bannermen, 9,965 rural bannermen, and 2,200 floating bannermen. Figure 8 contrasts the proportion of individuals disappearing from the data among different population categories, differentiating those who disappear because of missing registers and those, largely metropolitan bannermen, who disappear because of incomplete registration.

Accordingly, event-history analysis of the various types of exits using such flag variables as NEXT_DIE, NEXT_MARRY, and NEXT_REMARRY should normally be restricted to the one-year intervals where two consecutive triennial registers are both included in the CMGPD-SC.

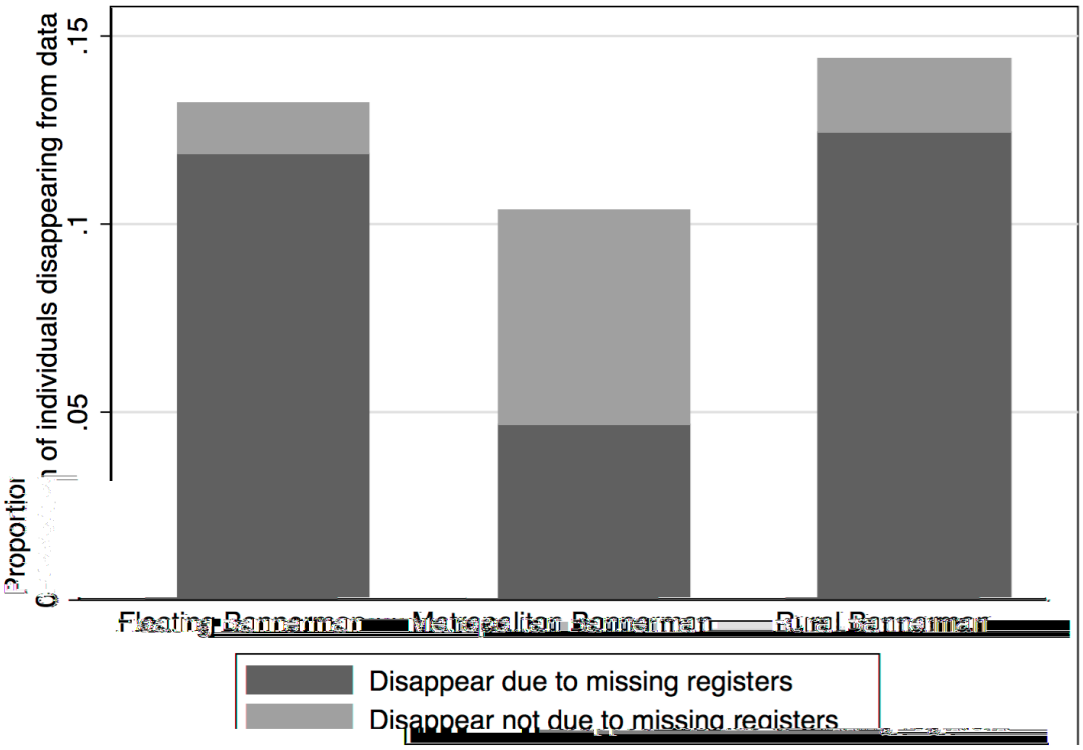


Figure 8 Proportion of individuals disappearing from the data

When multiple, consecutive registers are missing, the CMGPD-SC also omits individuals who first appeared in one of the missing registers and exited in a later missing register. Thus, for example, if 1874 and 1878 were available for a population but 1875, 1876, and 1877 were missing, a boy who appeared in 1875, was listed again in 1876, and whose death was recorded in 1877 would not be in the CMGPD-SC.

Similarly, a wife who was first listed in 1875 and whose death was listed in 1876 would not be in the CMGPD-SC.

2.C.III Bias in Timing and Occurrence of Events

The metropolitan and rural banner registers identify the year during which an event of interest occurred, not the precise date. For such exits the flag variables DIED, MARRIED_OUT, and REMARRIED_OUT mean that the event occurred in the previous year while NEXT_DIE, NEXT_MARRY and NEXT_REMARRY mean that they will occur in the following year. Flag variables are only set to 1 if the occurrence of the event is annotated in the current or subsequent register.

Marriages into a household, unlike such exits as death or out-marriage, are not specifically annotated in the original registers, thus relevant variables are constructed by the coders and software by comparing individual statuses across registers. If a man had no spouse listed for him in one register, but had a spouse listed with him in the following register, the software would set NEXT_MARRY to 1 for him in his observation in the first of the two registers. If he had been identified as a widower by the coders based on previous marital history, the software would set NEXT_REMARRY to 1 instead. The construction of NEXT_MARRY and NEXT_REMARRY accordingly differs for males and females. For females, NEXT_MARRY and NEXT_REMARRY are set to 1 only if the next available register positively identifies the female under consideration as marrying out via an annotation. For a male, the variables are set to 1 if he is unmarried or widowed in the current register but has a spouse listed with him in the next register in the CMGPD-SC, no matter how many of the intervening original registers are missing.

Variables such as NEXT_BOYS and NEXT_GIRLS that provide counts of births linked to a man or woman during a specific time frame are constructed by software based on year of birth calculated from their age in the first register that records them. They are not based on when registers first list a son or daughter, since in many cases sons or daughters appear for the first time in a register only when they were 5 or 6 years old (See 2.C.I above). Thus, for example, if a son first appears in 1876 and is listed as 5 *sui*, corresponding to a birth year of 1872, NEXT_BOYS would be set to 1 in 1871 for the father and mother. If no other births attributed to the mother and father took place between 1873 and 1874, NEXT_BOYS would be zero for them in 1873. NEXT_BOYS and NEXT_GIRLS, in other words, may not correspond to the actual presence or absence of sons or daughters in the subsequent register.

2.C.IV Covera

2.D Possible Applications

The CMGPD-SC is particularly valuable for studies of the social organization and demography of China, and to historical demography in general. The CMGPD-SC will contribute to illuminating relationships among socioeconomic stratification, household organization, and demographic behavior in historical China (Lee and Campbell 1997). Alongside its Liaoning counterpart, the CMGPD-SC can also be used for international comparisons of the role of household organization in modulating mortality and fertility responses to economic stress in past times (Bengtsson, Campbell and Lee 2004, Tsuya, et al. 2010).[18]

This user guide suggests below how the analysis of specific outcomes is especially likely to contribute to existing debates and controversies. We begin with a demonstration of the potential of the data to study the determinants of health and mortality since such applications are most clearly of direct and contemporary relevance. We then go on to demonstrate their potential utility for advancing our understanding of other phenomena in the social and behavioral sciences.[19]

2.D.I Mortality

The potential in the CMGPD-SC for analysis of mortality has yet to be exploited. Since the data provide complete life histories, they are a natural candidate for examinations of the influence of early life conditions on mortality later in life such

[18] Results from comparative studies involving the CMGPD-LN challenge longstanding orthodoxies about differences between Europe and Asia in the role of the family in responding to economic stress (Bengtsson, Campbell and Lee 2004, Tsuya, et al. 2010).
[19] Users are advised to refer to the CMGPD-LN User Guide (31-2) for a comprehensive review of previous CMGPD-LN-based work along these lines.

as those by Bengtsson and Lindstrom (2000, 2003) and Costa (2000).

Thr rich ert of

In addition, the detailed data on socioeconomic attainment and demographic behavior across and within generations makes it a potentially important source for studying how socially embedded demographic processes interact to shape population composition. With longitudinal information transcribed from repeated cross-sectional registers, the CMGPD-SC is superior to most contemporary data sources by allowing the direct measure of interactions between socioeconomic mobility, demographic differentials, and population composition. Shuangcheng population registration started shortly after the settlement of the banner population. This feature facilitates observing the evolution of socioeconomic inequality from an initial state of limited inequality in a frontier population.

A further unique advantage of the CMGPD-SC is the availability of data on wealth in the form of household landholding recorded at multiple points in time. This unusual feature renders the CMGPD-SC an exceptional source for studying inter- and intra-generational transmission and mobility of land, the most important form of wealth in historical China and elsewhere. Thus, the CMGPD-SC is likely to fill an important gap in the existing literature along this line. In a pioneering work using the land register data of the CMGPD-SC, Chen, Lee, and Campbell (2011) studied land stratification in Shuangcheng, focusing on the roles played by state institutions and individual agency. There is evidence for the persistence of the pattern of land distribution originally created by the state policy during the early years of the settlement. In fact, even after the related institutions were abolished, such patterns were largely maintained in spite of later land accumulation. In addition, the effects of social-economic status on the distribution of land differ between allocated and self-cultivated because of the different state regulations. Further exploitation of the land register data in the CMGPD-SC is expected to be especially fruitful.

2.D.IV Family, Kinship, and Community

One of the most remarkable features of the CMGPD-SC, alongside the CMGPD-LN, is the simultaneous detail on kin networks and communities. The simultaneous presence in the database of kin groups distributed by communities, and communities populated by different kin groups, allows for the disentanglement and measurement of kinship and community effects on individual demographic and social outcomes, as well as their interactions. This is likely to be one of the most fruitful areas of inquiry with these data. Even when understanding the interaction of kinship and community in shaping individual outcomes in a preindustrial population is not an end in itself, these data may be valuable as a testing ground for relevant estimation techniques.

2.D.V Institutional Context

Also still unexplored in the CMGPD-SC are the implications of institutional and social origin contexts for such outcomes as marriage, reproduction, morbidity, mortality and migration, or their implications for contexts such as family and descent line. As described at the beginning of this user guide, the Shuangcheng bannermen population is composed of three major groups defined by both place of origin and official immigrant status, i.e. metropolitan bannermen from Beijing, rural bannermen from elsewhere in Northeast China, and floating bannermen. The differences between categories and contexts were often quite stark. One of the most salient differences among these groups is the entitlement to land with metropolitan and rural bannermen

CMGPD-

3 CMGPD-SC Variables

We organized the variables of the CMGPD-SC into five major categories: basic variables, analytic variables, identifier variables, spatial variables, and property variables. Basic variables mainly cover demographic characteristics and events. Analytical variables are constructs suitable for event history analysis, fertility study, social stratification and mobility, among other things. The identifier variables are constructed by specific computer programs to facilitate dataset management. The meaning of spatial and property variables is straightforward.

Users will need permission to access the last two categories of variables, as well as individual names. In fact, these variables should be limited to use by investigators who have a verified affiliation with a research institution and who will commit to using the data only for quantitative, aggregated analysis. Names, spatial and property variables will be released in a separate data file, with access restricted to those who enter into an agreement with ICPSR as to their use. Our desire is to prevent the CMGPD-SC from being used for genealogy, whether by individuals or commercial enterprises, and from being used for historical or biographic research on specific, named individuals. Because ICPSR is intended to support academic research and we are carrying out academic research, we are not in a position to provide support for anyone engaged in such applications. The original registers from which the CMGPD-SC was transcribed are better suited to such applications.

A number of variables that are included in the CMGPD-LN are not available in the CMGPD-SC because they are only relevant to the CMGPD-LN context, or not provided in the CMGPD-SC raw data. REGION and DISTRICT, geographic variables that identify regions and administrative units within Liaoning, are not applicable in Shuangcheng, which itself is just a county. Since the Shuangcheng banner population did not have the organization of *zu*, ZU_ZHANG and ZU_SEQ are not applicable. Due to important institutional disparities between the banner populations covered by the CMGPD-LN and CMGPD-SC, many positions and statuses are not meaningful for the Shuangcheng banner population, including OLD, ARTISAN, BAIZONG, BAOYANG, EXPELLED, GAO_LI, MAN_ZHOU, SERVICE_DING, QIAN_ZONG, TOU_CONG, and ZHI_SHI_REN. See the CMGPD-LN User Guide for detailed descriptions of these variables (Lee, Campbell and Chen 2010).

3.A Handling of Missing Values

We distinguish two forms of missing values: regular and structural. We follow the same standard for defining these variables as in the CMGPD-LN release.

We use -99 to identify values that are missing in the sense that observations of this type in the original data normally record the information contained in this variable, but that in this particular observation, there was no such information. A common example is AGE_IN_SUI. Most observations in the original data record an age, but some do not. Observations in which the individual is annotated as having died or

otherwise exited since the last register are especially likely to omit an age. In those observations where there is no age recorded in the original register, AGE_IN_SUI is set to -99 in the CMGPD-SC release. While AGE_IN_SUI provides the most examples of such missing values, there are missing values for other variables as well. While these mostly reflect clerical errors in the original registers that led to an omission, in some cases they may reflect that something was written, but was illegible.

We use -98 to identify values that missing because observations of this type in the original data normally did not record the information contained in the variable. One example is the various variables for administrative status in observations of women. Since the original data do not normally record administrative status for women, these variables have been set to -98 for all females.

3.B Basic Variables

Coders transcribe or assign the values of basic variables directly from the contents of the original population registers. In some cases, they are flag variables identifying whether or not a particular annotation was present in the original record, or whether the original record indicated a particular status for a person. Some basic variables are constructed from the following information on the record pages: relationship, place of residence, age (in *sui*), record of demographic events, administrative status indicator, and identification of the head of the household (See Figure 2 above).

compiled. These values are transcribed and converted by the coders from the imperial reign year recorded on the title page of each register. Figure 9 shows the distribution of observations by calendar year.

Figure 9 Numbers of Observations by Calendar Year

NAME

NAME is the *hanyu pinyin* for the original name recorded in the register. This variable is not included in the public release of the CMGPD-SC, but will be included in a restricted release that will be available from ICPSR to users who complete

(i.e. number, and o and y), resulting in 97 distinct basic relationship types.

Table 2 Interpretation of codes in RELATIONSHIP

Code	Relationship
e	Ego or household head
w	

	(4.60)	(9.43)	(5.77)	(8.59)
Household head	**67,924**	**418,508**	**66,766**	**553,198**
	(49.31)	**(38.73)**	**(51.96)**	**(41.07)**
One generation down	57,695	419,569	46,367	523,631
	(41.88)	(38.83)	(36.08)	(38.88)
Two generation down	5,266	122,706	7,521	135,493
	(3.82)	(11.36)	(5.85)	(10.06)
Three generation down	32	7,676	255	7,963
	(0.02)	(0.71)	(0.20)	(0.59)
Four generation down	0	105	2	107
	(0.00)	(0.01)	(0.00)	(0.01)
Five generation down	0	2	0	2
	(0.00)	(0.00)	(0.00)	(0.00)
Total	137,750	1,080,577	128,499	1,346,826

Source: CMGPD-

AGE_IN_SUI

AGE_IN_SUI is the direct transcription of the age information from the original register. Ages recorded in the Shuangcheng registers are always written in *sui*, a traditional way to calculate age in China. A person is aged 1 *sui* at birth and is one year older after each lunar new year. On average, an age measured in *sui* is 1.5 years older than an age reckoned in the Western method. Since birthdate is not recorded in the SC registers, there is no means of directly calculating an age in Western years. To facilitate comparison with results from elsewhere in which ages are in Western years, in our own analysis we generally define age groups with the initial and final year offset one by year. For example, to produce something comparable to Western ages 5-9, we typically use the age range 6-10 *sui*.

In the CMGPD-SC data, for a non-trivial portion of the observations (n=61,681, 4.58%), AGE_IN_SUI is missing and coded as -99. This is normally the case when an individual is listed as dead or otherwise not present. Users should not include these observations into analysis. Normally selecting observations for AGE_IN_SUI not equal to -99 and PRESENT equal to 1 will eliminate these observations. In the CMGPD-SC, there are 1,264,671 observations with non-missing AGE_IN_SUI and PRESENT equal to 1 for 103,837 individuals.

When using AGE_IN_SUI, users should also note that there are occasional

Lee, Campbell, and Chen (2010), this problem can be addressed by creating a calculated age in each register based on the BIRTHYEAR calculated from the earliest observation of that individual. (See below for the description of BIRTHYEAR.)

Figure 10 shows the age distribution for the CMGPD-SC observations. A comparison of Figure 10 with the age distribution for the CMGPD-LN observations suggests that the underreporting of girls aged less than 15 *sui* is appreciably less severe in CMGPD-SC. There is also evidence that age heaping is less severe in the CMGPD-SC than in the CMGPD-LN. However, the underreporting of infants and young children remains appreciable. It is worth noting that among the three bannermen groups, metropolitan bannermen seems to have been best recorded in the registers in all respects.

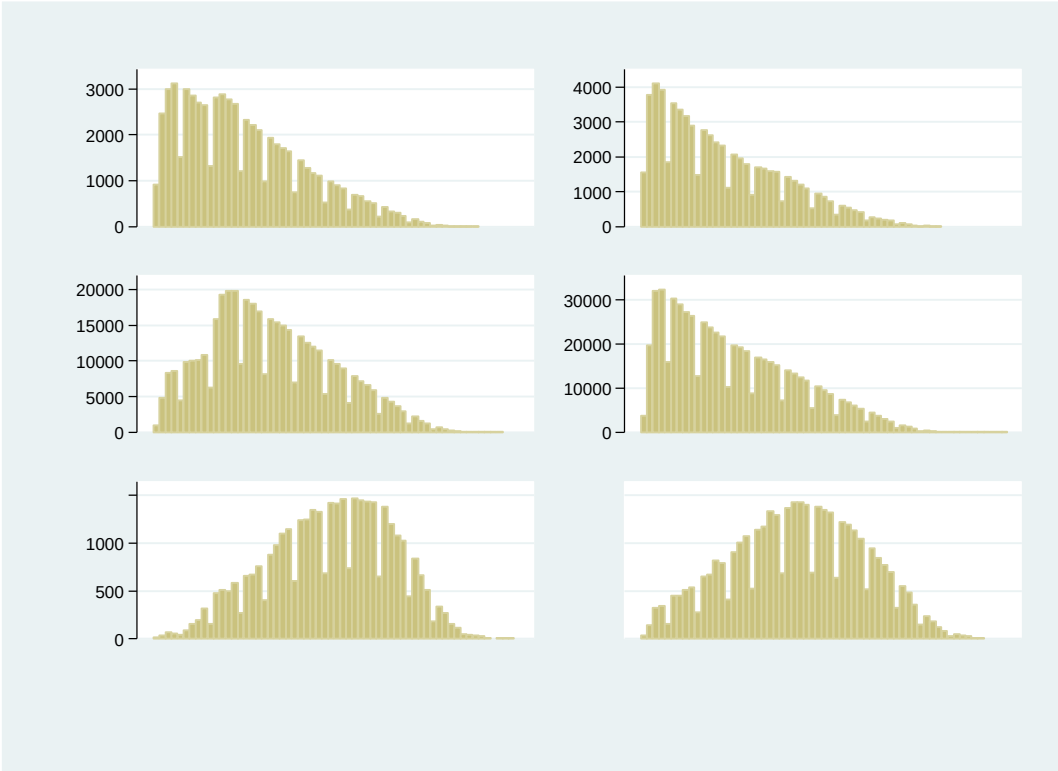


Figure 10 Age Distribution by SEX and Banner

Source: CMGPD-SC, 1866-1913.
Note: Only observations with valid age (1-110 *sui*) and PRESENT=1 are included.

Because of problems with the original data, some men appear to survive to absurdly old ages. This phenomenon is discussed in detail in **Error! Reference source not ound.** because of its impact on the study of mortality at advanced ages. In our own analysis, we generally exclude observations of individuals aged 76 *sui* or higher on the assumption that a large share of them were actually already dead, but still being recorded. An even more conservative approach would be to exclude all of the records of anyone who survives to an absurdly advanced age. For example, `BYSORT PERSON_ID (AGE_IN_SUI): drop if AGE_IN_SUI[_N-1]>= 91` would eliminate all the observations of any individual who appeared to reach 91 *sui* or higher by their last available record.

BIRTHYEAR

BIRTHYEAR is a generated variable calculated from the age recorded in the original registers. It is calculated as YEAR-AGE_IN_SUI+1. This variable provides an alternative basis for the calculation of age, and at least an approximation of age reckoned according to the Western standard. To produce a consistent age variable without any of the discrepancies in recorded AGE_IN_SUI discussed above, BIRTHYEAR can be copied forward from an individual's first record to their later records, and then subtracted from YEAR. BIRTHYEAR ranges from 1770 to 1913 in the CMGPD-SC.

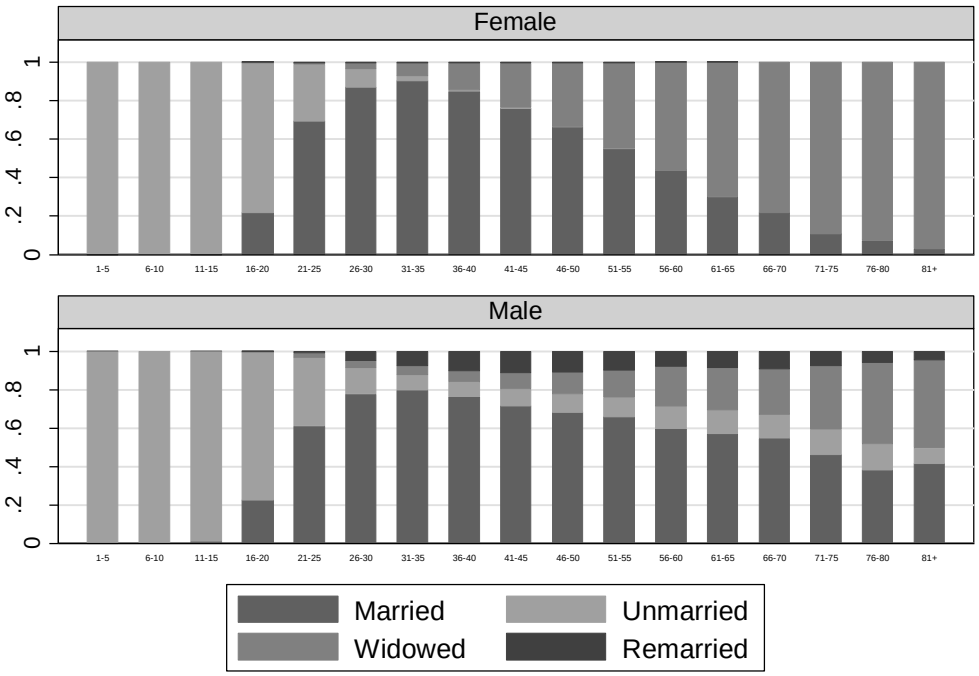
ETHNICITY

ETHNICITY is the ethnicity of the household head. Of 1,343,893 observations with valid information on ethnicity, 44.74 percent are Manchu,

User Guide for an extended discussion on of the procedure. Figure 11 provides age-specific marital status by sex for the three CMGPD-SC banner populations.

Marital status is coded inconsistently for 177 observations of 155 individuals, in the sense that their marital status is coded as unmarried after being coded as married or widowed in a previous year. Each of these individuals has on average 20 observations. Users should exclude these observations from the analysis or generate their own variable for marital status as needed.

(a) Metropolitan bannermen



(b) Rural bannermen

FATHER_NAME is the *hanyu pinyin* for the name
is recorded in some Qing and almost all Republican registers (See Figure 2). This variable is not included in the public release of the CMGPD-SC, but will be available in a restricted release from ICPSR to users who sign a confidentiality agreement. In principle, we can identify siblings across households by FATHER_NAME coupled with other necessary information. FATHER_NAME is particularly useful for assigning FATHER_ID[24] to those who were recorded as household heads in the earliest available population registers compiled in 1866.

While father s name is commonly recorded in the CMGPD-LN, it only occurs for a small percentage of the CMGPD-SC. Out of 3,087 household heads in the 1866 Shuangcheng registers (coded as e' on RELATIONSHIP), 492 (16%) have father's given name recorded. Also, over 90 percent of them are *tunding* affiliated with Plain Blue banner (n=428). These 428 individuals accounted for about 85 percent of all Plain Blue banner *tunding*. Further examination revealed that they were all from regions in Liaoning, including Jinzhou, Fuzhou, Fenghuangcheng, Xiongyue, and Kaiyuan. In particular, 63.7 percent of them migrated from Jinzhou near present-day Dalian.

GRANDFATHER_NAME

GRANDFATHER_NAME is the *hanyu pinyin* for the
grandfather which is recorded in some Qing and almost all Republican records in the register. This variable is not included in the public release of the CMGPD-SC, but will be available in a restricted release from ICPSR to users who sign a confidentiality agreement. Among 3,087 household heads (coded as e' on RELATIONSHIP) in the 1866 Shuangcheng registers, 427 (14%) have grandfather's given name recorded. Again, almost all these cases are *tunding*.

FATHER_POSITION

FATHER_POSITION is the numeric code of the original statuses recorded for household head household head The codes are assigned by coders when transcribing the information from the household registers. This variable is not included in the current release of the CMGPD-SC but is available upon request.

GRANDFATHER_POSITION

GRANDFATHER_POSITION is the numeric code of the original statuses recorded for household head
codes are assigned by coders when transcribing the information from the household registers. This variable is not included in the current release of the CMGPD-SC but is available upon request.

[24] See FATHER_ID below.

3.B.II Events

In the Eight Banner population registers, individual exits such as death, out-marriage, out-adoption, and absconded, are explicitly recorded in each
explain the disappearance of the observed individual since the last register. For example, in the record page shown in Figure 2, one individual is annotated as having died during the past year. Specifically, for metropolitan and rural banner registers, the annotations refer to occurrences in the last year, while for floating bannermen the annotations refer to occurrences within the three years since the previous register. Events associated with the appearance of an individual such as birth, marriage, and adoption were generally inferred by the coders and were not explicitly recorded in the original registers. Based on the original records of vital events, and the coders' inferences based on changes between registers, we created the following variables to capture the most important demographic events in this population.

Because the exits recorded in the original registers refer to events that have already taken place, the corresponding flag variables DIED, MARRIED_OUT, REMARRIED_OUT and ABSCONDED below normally should not be used as dependent or outcome variables in an analysis. Additionally, observations in which an individual is annotated as exiting commonly omit their age, thus in most observations in which any of these exit flag variables are 1, for true, AGE_IN_SUI is missing, -99.

Flag variables such as NEXT_DIE that the software constructs from these original exit variables specify the occurrence of the event in the time between the current and next register, and are preferable as dependent variables. We include the original flag variables DIED, MARRIED_OUT, REMARRIED_OUT and ABSCONDED so that the dataset maintains as much of the original data as possible. They are useful mainly for the construction of new outcome variables if the user is dissatisfied with the ones provided.

EVENT_1

EVENT_2

EVENT_1 and EVENT_2 are numeric values as entered by the coders to indicate the occurrence of specific types of entrances and exits from registers. Most of these values are based on annotations in the registers that indicate the occurrence of specific events in last year for metropolitan and rural bannermen and in the past three years for floating bannermen. Table 6 summarizes the annotations or other events corresponding to each of the values of EVENT_1 and EVENT_2. Exits since the last register were particularly likely to be coded in the original registers. For example, if an individual recorded in an observation had died in last year, their entry in the register included the annotation *wang gu* (dead). The other most commonly recorded reasons for exit were *chu jia* (out-marriage) for daughters, *gai jia* (remarriage) for widows, and *chu tao* (absconded) for males. In the case of birth (6), in-marriage (7), or new appearance (9), the values were assigned by coders for observations of individuals who were appearing in the registers for the first time, and reflected the coders' assessment of the likely reason for the entrance. They do not reflect an annotation in the original register. We have two variables for events because in some cases, coders entered two values for events, one for an entrance

that they inferred by comparison with the preceding register, and another for an exit recorded directly in the register.

Table 6 Explanation of EVENT_1 and EVENT_2 Codes

Event

Table 9 - and population category

	Banner Population					
	Metropolitan		*Rural*		*Floating*	
	No	Remarried-out	No	Remarried-out	No	Remarried-out
Female	66,755	111	443,468	492	41,262	36
Male	70,879	5	636,603		87,201	
Total	137,634	116	1080071	492	128,463	36

Source: CMGPD-SC, 1866-1913.

Again, according to **Error! Reference source not found.**, only a small proportion of .

Table 10 Number of Remarried-out Observation per Individual

Observations annotated as remarried-out	Banner Population			
	Metropolitan	*Rural*	*Floating*	Total
0	9,013	79,476	18,505	106,994
	(0.99)	(0.99)	(1.00)	(0.99)
1	60	405	34	499
	(0.01)	(0.01)	(0.00)	(0.00)
2	13	38	1	52
	(0.00)	(0.00)	(0.00)	(0.00)
>=3	3	3		6

	Metropolitan	*Rural*	*Floating*	Total
No	70,543	635,959	87,154	793,656
Absconded	341	644	47	1,032
Total	70,884	636,603	87,201	794,688

Source: CMGPD-SC, 1866-1913.

DIED

DIED is a dummy variable. The value 1 indicates that an individual was annotated in the original register as having died sometime since the previous register.

There are individuals who were annotated as dead in more than one register, and thus have values of 1 for DIED in more than one observation (Table 12). In most cases, this occurs because the record of a dead individual was copied forward into subsequent registers without being expunged. H
back to lif s (e.g. PERSON_ID= 00266287).

Table 12 Number of Annotated Death (*gu*) per Individual

	Banner Population			
#Died	*Metropolitan*	*Rural*	*Floating*	Total
0	6,485	60,421	12,392	79,298
	(0.71)	(0.76)	(0.67)	(0.74)
1	1,775	15,476	4,749	22,000
	(0.20)	(0.19)	(0.26)	(0.20)
>=2	829	4,025	1,399	6,253
	(0.09)	(0.05)	(0.08)	(0.06)
Total	9,089	79,922	18,540	107,551

Source: CMGPD-SC, 1866-1913.

PRESENT

PRESENT is a dummy variable generated by a specific computer program based on the vital demographic events recorded in the register. Generally, a person is counted as present if he/she is not annotated as having exited since the last register, i.e. DIED, MARRIED_OUT, REMARRIED_OUT, ABSCONDED are all 0, and no other form of exit is recorded in the register.

Table 13 Observations with Individual Annotated as Alive and Present

		Banner Population			
Present		*Metropolitan*	*Rural*	*Floating*	Total
	No	7,628	42,499	31,998	82,125
		(0.06)	(0.04)	(0.25)	(0.06)
	Yes	130,122	1,038,078	96,501	1,264,701
		(0.94)	(0.96)	(0.75)	(0.94)

Total	137,750	1,080,577	128,499	1,346,826

Source: CMGPD-SC, 1866-1913.

This variable is useful to define the population at risk when analyzing exiting events such as death, out-marriage, and etc. Moreover, based on this variable, an array of flag variables was created to identify the occurrence of a demographic event to the observed individual in next available register.

Users should keep in mind that since the floating banner registers are updated triennially instead of annually, when using the Next_* variable in analysis, the vital events of floating bannermen should occurred in the past three years instead of one.

3.B.III Administrative Statuses

The majority of the variables in this section, unless otherwise noted, are generated by software based on the contents of the coder's original transcription into pinyin of the administrative status recorded for an individual in a register. In the Shuangcheng

surname (Figure 2). These statuses are usually information relevant to the individual's relationship to the state, and typically consisted of salaried state positions, official ranks, honorary and exam titles, and/or disability, if any. Sometimes, information on ethnicity is also recorded for those whose ethnicity differed from the majority of the people in the register. Users can use these variables to analyze the causes and consequences of the socioeconomic and political status of an individual. Through additional manipulations, it is also possible to use these to construct variables measuring the characteristics of the members of an individual's household or kinship network.

Similar to the CMGPD-LN, administrative status variables are only valid for males and are set to missing (-98) for females. There are significant disparities between administrative statuses recorded on the Liaoning and Shuangcheng population registers , reflecting important institutional differences. Being administered by the Shengjing Imperial Household Agency, the CMGPD-LN population has a wide variety of positions under the banner, civilian, and local systems. However, the majority of the positions in CMGPD-SC are military titles exclusively under the banner system. Therefore, the distributions of administrative statues follow distinct patterns in the CMGPD-SC and CMGPD-LN.

POPUATION_CATEGORY

POPULATION_CATEGORY is the numeric code assigned to the registration category of the household, which appeared on the cover page of the household registers. There are three registration categories in the CMGPD-SC population: *jingqi, tunding, and fuding*. These registration categories differentiated the entitlement rights to land assigned to the sub-groups of the CMGPD-SC population. *Jingqi* or metropolitan bannermen were the highest-status category, each household owning 35 *shang* of state-allocated land. *Tunding* or rural bannermen were the middle-status category; each household owned 18.33 *shang* allocated land. *Fuding* or floating

bannermen were the lowest-status category; they had no entitlement rights to allocated land.

POSITION_CODE

POSITION_CODE is the numeric code assigned to the original administrative statuses recorded for some males. The codes were assigned by coders when transcribing the information from the household registers. These original statuses are used to generate the variables regarding status and position in the analytical release. Through POSITION_CODE, users can retrieve both the pinyin and the Chinese characters for the specific statuses. In CMGPD_SC, about 20.5 percent of the male observations have a recorded administrative status. For the remaining 79.5 percent of male observations that have no recorded administrative status, the value is set to -99.

DISABILITY_CODE

DISABILITY_CODE is a numeric code for disability. Coders assigned the values during entry. It can be used to automatically retrieve the pinyin for the disability recorded in the population registers from an accompanying data file, or to manually look up the original Chinese characters in the Codebook Appendix.

Since the Shuangcheng banner population did not provide labor to the state, information on disability was seldom recorded on the registers. Only 0.27 percent of the male observations had disability information, while for 99.72 percent of the male observations the value was set to -99.

NO_STATUS

NO_STATUS is a flag variable indicating that a person has no recorded administrative status in the register under consideration. The value 1 indicates that the individual had no recorded administrative status, and a 0 indicates that the individual was recorded as having administrative status. Most males had an administrative status. Since females were not normally eligible for an administrative status, this variable is coded as a missing (-98) for them. Most males did not have an administrative status recorded until they reached 18 *sui*.

Table 14 Administrative Status for Male Observations, 18-60 sui

	Banner Population			
	Metropolitan	*Rural*	*Floating*	Total
Any Status	24,318	67,617	5,363	97,298
	(0.69)	(0.19)	(0.13)	(0.23)
No Status	10,903	280,275	36,730	327,908
	(0.31)	(0.81)	(0.87)	(0.77)
Total	35,221	347,892	42,093	425,206

Source: CMGPD-SC, 1866-1913.
Note: Only observations with valid age (1-110 *sui*) and PRESENT=1 are included;

Table 16 Age at Retirement for Individuals Ever Annotated as Retired *(tui*)

Age	Banner Population			
	Metropolitan	*Rural*	*Floating*	Total
-99	55	274	1	330
	(0.38)	(0.42)	(0.25)	(0.41)
1-10	0	1	0	1
	0.00	(0.00)	0.00	(0.00)
11-20	2	8	0	10
	(0.01)	(0.01)	0.00	(0.01)
21-30	15	44	0	59
	(0.10)	(0.07)	0.00	(0.07)
31-40	28	116	0	144
	(0.19)	(0.18)	0.00	(0.18)
41-50	26	107	2	135
	(0.18)	(0.16)	(0.50)	(0.17)
51-60	18	63	1	82
	(0.12)	(0.10)	(0.25)	(0.10)
61-70	1	32	0	33
	(0.01)	(0.05)	0.00	(0.04)
71+	0	14	0	14
	0.00	(0.02)	0.00	(0.02)
Total	145	659	4	808

Source: CMGPD-SC, 1866-1913.
Note: Only observations with valid age (1-110 sui) and PRESENT=1 are included;

HAS_POSITION

HAS_POSITION is a flag variable indicating that a male had a salaried position at the time the register was produced. Since the Shuangcheng banner administration followed the model of the Eight Banner garrison, the majority of the salaried positions are military positions divided into two categories: officials and soldiers. In addition, village heads in Shuangcheng also received state stipends. Salaries mainly came in the form of money. Since salaried positions were available only to males, the variable is set to missing (-98) for females. The calculation based on male observations indicates that the proportion of observations annotated as holding a salaried position is much higher for metropolitan bannermen than for either rural bannermen or floating bannermen (See Table 17). Furthermore, 7.76% of metropolitan bannermen males had *ever* held a salaried position, while the proportions are only 1.87% and 0.09% for rural bannermen and floating bannermen, respectively.

Table 17 Salaried Position Status for Male Observations, 18-60 sui

Held salaried position	Banner Population			Total
	Metropolitan	*Rural*	*Floating*	
No	31,231	340,985	42,074	414,290
	(0.89)	(0.98)	(1.00)	(0.97)
Yes	3,990	6,907	19	

Total	35,221	347,892	42,093	425,206

Source: CMGPD-SC, 1866-1913.
Note: Only observations with valid age (1-110 *sui*) and PRESENT=1 are included;

RANK

status. The numeric value is transcribed from the original records in the household registers. The bureaucratic rank in the Qing ranged from 1, the highest, to 9, the lowest. Moreover, these ranks not only applied to official titles but also were used as affixes to honorific titles, indicating the degree of honor.

HONORIFIC

honorific titles. The honorific titles in Shuangcheng include *dingdai* (official hat ornament), *lanling* , *hualing* (ornament of offi , *jungong* (award to military merit), *gaofeng* (conferred by imperial mandate), and etc.

EXAMINATION

EXAMINATION is a flag variable indicating whether a male held a government student title or degree achieved under the government exam system, including both civil service and military exams. The civil service examination titles in Shuangcheng included *wentong, shengyuan, juren, jinshi*. The military examination titles included *wutong, wusheng,* and *wuju*.

These titles were either purchased or earned by taking an exam. This variable can be used to measure the individual and his family's educational achievement or investment in education.

JUANNA

JUANNA
purchased. Beginning in the late seventeenth century, the Qing government sold various titles to collect money to subsidize military expenditures. This practice continued until the end of the Qing dynasty and became especially important after the mid-nineteenth century. The titles sold include real official titles, honorific titles, and exam degrees. This variable can be used as an indicator of the individual and his status, since purchasing a title required a considerable amount of money.

PURCHASED_TITLE

FATHER_ALIVE

married females, or husband s father for married or widowed women. The value 1 indicates that the father (or father-in-law) of the observed person is alive and living in the household in the current register. The value 0 indicates that the father (or father-in-law) of the observed person was either dead or not living in the household in the current register. The proportion of metropolitan and rural banner males who had ever lived with their fathers alive for the period covered by CMGPD-SC is slightly over 80% if the calculation is limited to those who were first registered at age 50 or younger. The corresponding proportion for floating banner males is only about 66%.

Figure 14 shows age-specific proportions of living with father alive for males.

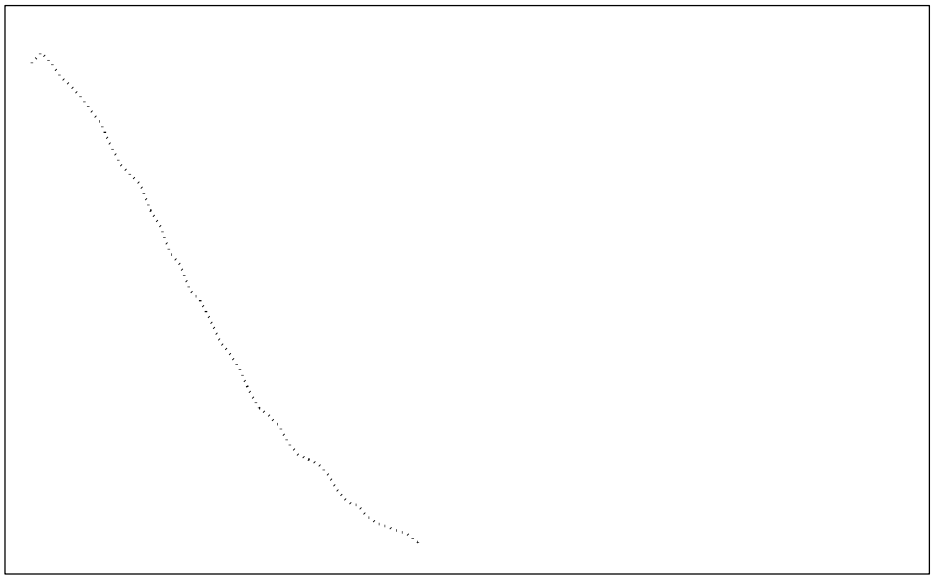


Figure 14 Age-specific Proportion of Males Living with Father
Source: CMGPD-SC, 1866-1913.
Note: Restricted to males with valid age (in *sui*) and PRESENT=1.

MOTHER_ALIVE

MOTHER_ALIVE is a flag variable that refers to mother for males and never-married females, or mother-in-law for married or widowed women. The value 1 indicates that the mother (or mother-in-law) of the observed person was alive and living in the household in the current register. The value 0 indicates that the mother (or mother-in-law) of the observed person was either dead or not living in the household in the current register. The variable can be treated as an indicator of the survival of the mother. Nearly 90% of metropolitan and rural banner males had lived with their

mothers alive for the period covered by CMGPD-SC if the calculation is limited to those who were first registered at age 50 or younger. The corresponding proportion for floating banner males is about 71%. **Figure 15** shows age-specific proportions of males living with mother.

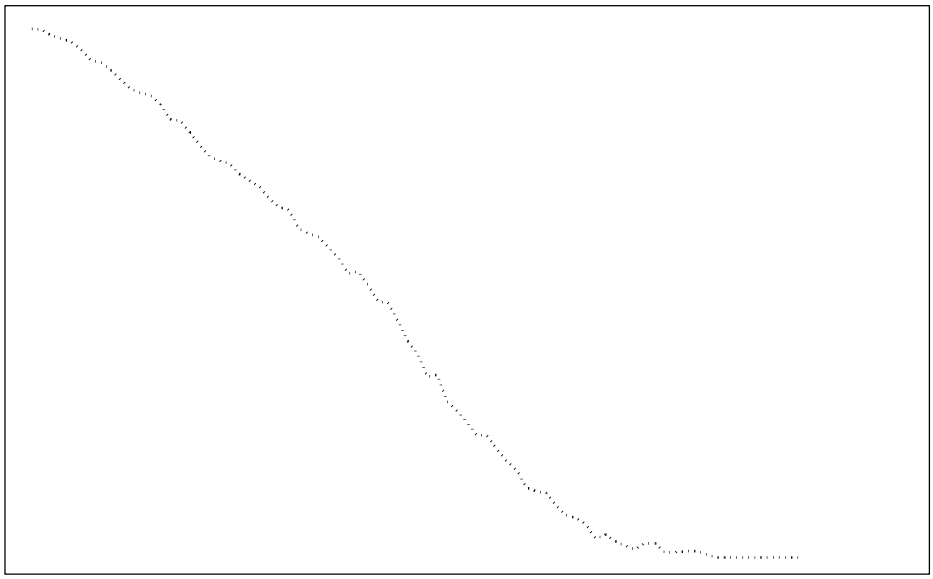


Figure 15 Age-specific Proportion of Males Living with Mother
Source: CMGPD-SC, 1866-1913.
Note: Restricted to males with valid age (in *sui*) and PRESENT=1;

SON_COUNT

SON_COUNT is the number of sons born to the observed individual up to the year of the current register. The values were generated by software based on record linkage. It is not based on the presence of sons recorded in the current register, but rather on a separate calculation of the number of males whose FATHER_ID or MOTHER_ID was this person's PERSON_ID, and who had a calculated year of birth up to or including the current register year. This variable is time-varying. SON_COUNT may differ from the number of sons apparent in the register because it will include sons who appear in later registers, but who were born before the current register. This

history into an analysis. This variable likely underestimates the actual number of sons born to an individual because, as discussed elsewhere, many boys who died in infancy or early childhood were never recorded in the registers. Accordingly, SON_COUNT should be thought of as a count of sons who survived long enough to be registered. Figure 16 shows the average number of sons by age, while Table 18 shows the distribution of the number of boys ever born.

Figure 16 Mean SON_COUNT by Age of Parent

Source: CMGPD-SC, 1866-1913.
Note: Restricted to those with a valid age (in *sui*) and PRESENT=1.

Table 18 Distribution of SON_COUNT

	Banner Population [redacted]
#Sons ever born	

DAUGHTER_COUNT is the number of daughters born to the observed individual up to the year of the current register. Like SON_COUNT, this variable is also time-varying. Please see the discussion of SON_COUNT. Because many daughters were never registered, this variable undoubtedly underestimates the actual number of daughters.

Figure 17 shows the average number of daughters by age, while Table 19 shows the distribution of the number of daughters ever born.

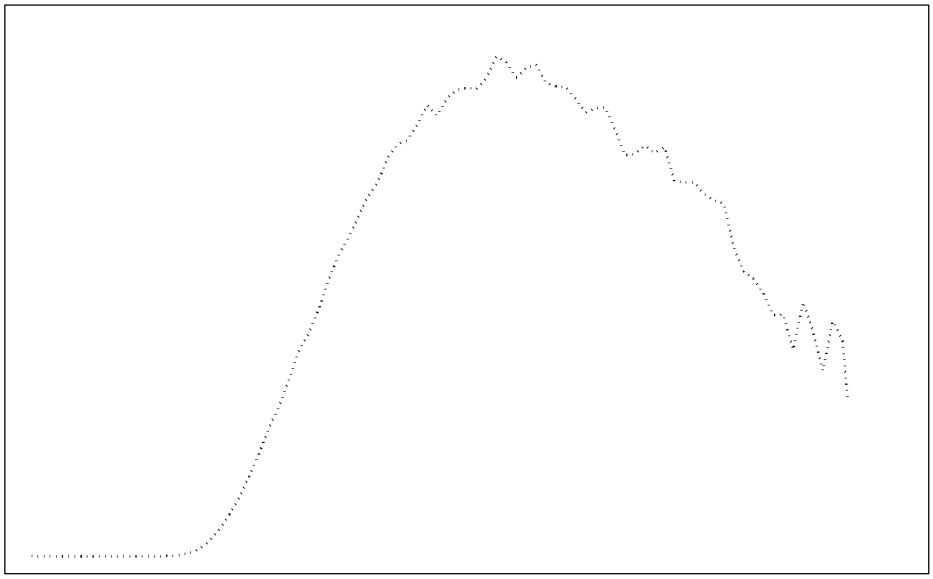


Figure 17 Mean DAUGHTER_COUNT by Age of Parent

Source: CMGPD-SC, 1866-1913.
Note: Restricted to those with a valid age (in *sui*) and PRESENT=1.

Table 19 Distribution of DAUGHTER_COUNT

	Banner Population			
#Daughters born	*Metropolitan*	*Rural*	*Floating*	Total
0	4,225	46,636	9,681	60,542
	(0.84)	(0.88)	(0.93)	(0.88)
1	511	4,788	644	5,943
	(0.10)	(0.09)	(0.06)	(0.09)
2	187	1,268	68	1,523
	(0.04)	(0.02)	(0.01)	(0.02)
3	60	286	7	353
	(0.01)	(0.01)	(0.00)	(0.01)
4	22	68	2	92
	(0.00)	(0.00)	(0.00)	(0.00)
510		6	0	16
	(0.00)	(0.00)	0.00	(0.00)
>=6	2	0	0	2
	(0.00)	0.00	0.00	0.00

	5,017	53,052	10,402	68,471

Source: CMGPD-SC, 1866-1913.
Note: Restricted to those who are either married or remarried or widowed;

OTHER KIN COUNT VARIABLES. For males and never-married daughters, BROTHER_COUNT, for example, is the number of male siblings living in the household in the current register. For married and widowed women, it is the number of husband's brothers living in the same household. Note that all such count variables are time-varying. The same rule also applies to SISTER_COUNT, MALE_COUSIN_COUNT, FEMALE_COUSIN_COUNT, UNCLE_COUNT, and AUNT_COUNT. Also note that only cousins, uncles, and aunts along the paternal line are counted. Note that our discussion will mainly focus on males.

BROTHER_COUNT

For males and never-married daughters, BROTHER_COUNT is the number of male siblings living in the household in the current register. For married and widowed women, it is the number of husband's brothers living in the same household. The values were generated by a computer program based on the presence of other men in the household with the same FATHER_ID. Figure 18 shows the average number of brothers of males and unmarried daughters living in the same household by age, while Table 20 shows the distribution of the number of brothers of males and unmarried daughters ever registered in the same household..

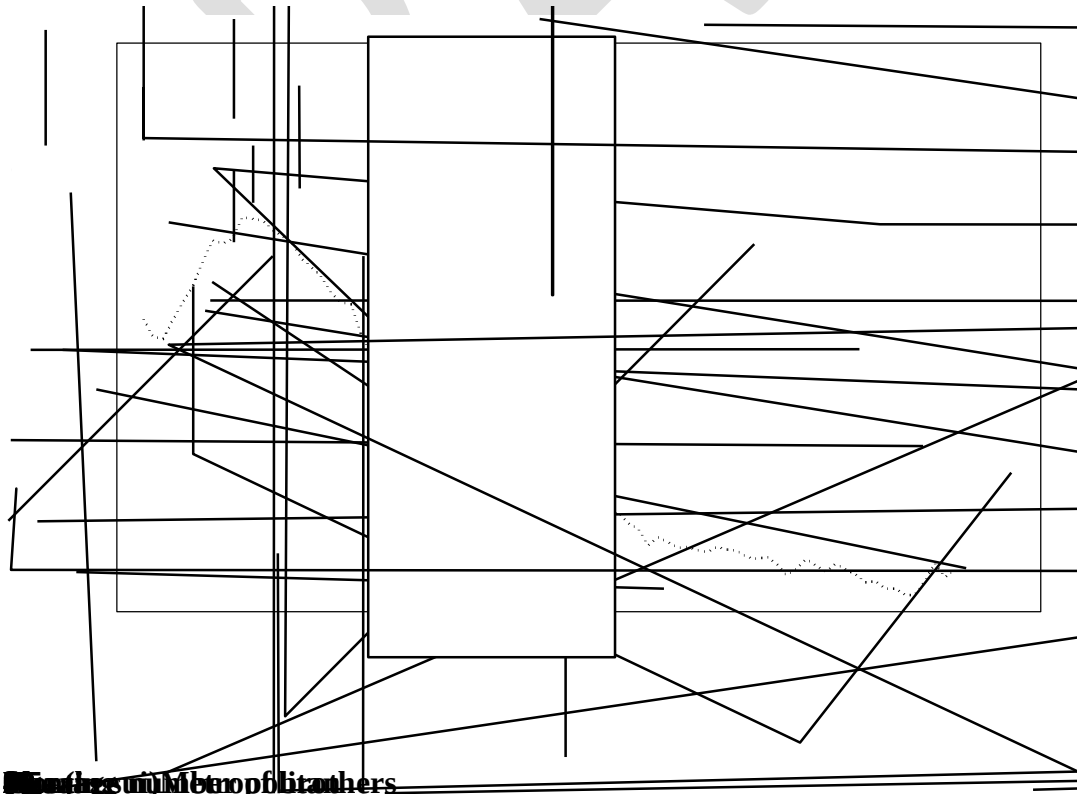


Figure 18 Average BROTHER_COUNT by Age

Source: CMGPD-SC, 1866-1913.
Note: Restricted to males with valid age (in *sui*) and PRESENT=1.

Table 20 Number of BROTHER_COUNT

#Brothers ever had	Banner Population			
	Metropolitan	*Rural*	*Floating*	Total
0	1,622	12,783	3,447	17,852
	(0.25)	(0.25)	(0.34)	(0.26)
1	2,067	15,982	2,941	20,990
	(0.31)	(0.31)	(0.29)	(0.31)
2	1,541	11,998	2,030	15,569
	(0.23)	(0.23)	(0.20)	(0.23)
3	824	6,866	1,014	8,704
	(0.13)	(0.13)	(0.10)	(0.13)
4	340	2,918	452	3,710
	(0.05)	(0.06)	(0.05)	(0.05)
5	114	1,124	105	1,343
	(0.02)	(0.02)	(0.01)	(0.02)
>=6	66	512	26	604
	(0.01)	(0.01)	(0.00)	(0.01)
	6,574	52,183	10,015	68,772

Source: CMGPD-SC, 1866-1913.
Note: 2.85 52.68 Tm[(1)] TJETQq90.024 49.68 382.22(a)4(n)-6(d)1

SISTTJETJR_C 38OUNT

For ma.68 and nTmver
siblings living in the household in the current register. For married and widowed women, it is the number of sisters-in-law including husband's brothers' wiv68and unmarried sisters

19 shows the amverage number ofsisters of m4 49.68 and unmarried daughters living in the household by age, while Table 21 shows the distribution of the number of sisters registered in the same household.

Figure 19 Average SISTER_COUNT by Age

Source: CMGPD-SC, 1866-1913.
Note: Restricted to males with valid age (in *sui*) and PRESENT=1.

Table 21 Number of SISTER_COUNT

	Banner Population			
#Sisters ever had	*Metropolitan*	*Tunding*	*Floating*	Total
0	3,219	38,697	8,986	50,902
	(0.50)	(0.75)	(0.94)	(0.75)
1	1,840	9,779	503	12,122
	(0.29)	(0.19)	(0.05)	(0.18)
2	907	2,652	56	3,615
	(0.14)	(0.05)	(0.01)	(0.05)
3	349	553	8	910
	(0.05)	(0.01)	(0.00)	[redacted]
4				

MALE_COUSIN_COUNT

For males and never-married daughters, MALE_COUSIN_COUNT is the number of male paternal cousins living in the household in the current register. For married and widowed women, it is the number of husband's male paternal cousins living in the same household. The values were generated by a computer program based on the presence of other men in the household who had the same GRANDFATHER_ID. Figure 20 shows the average number of male cousins of males and unmarried daughters living in the household by age, while Table 22 shows the distribution of the number of male cousins of males and unmarried daughters ever registered in the same household.

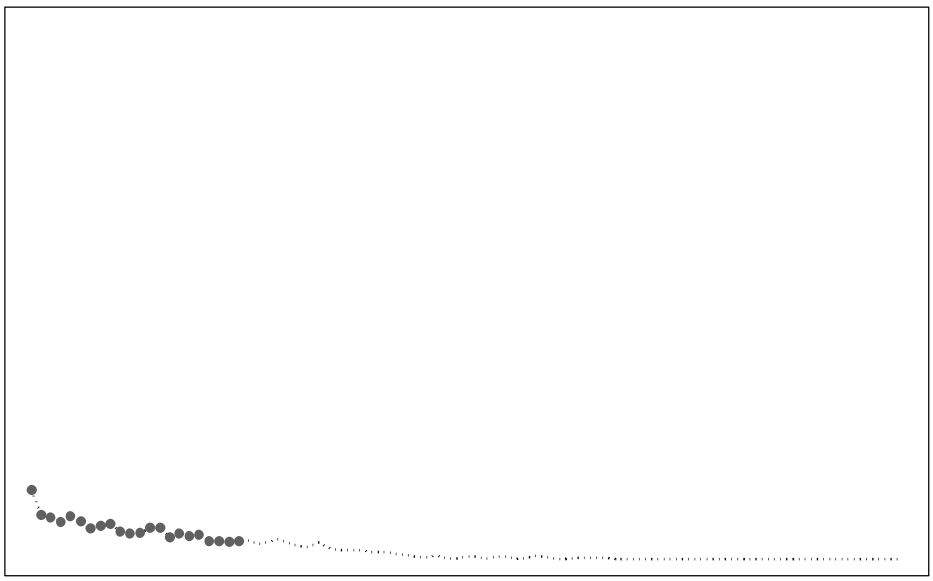


Figure 20 Average MALE_COUSIN_COUNT by Age

Source: CMGPD-SC, 1866-1913.
Note: Restricted to males with valid age (in *sui*) and PRESENT=1.

Table 22 Number of MALE_COUSIN_COUNT

#Male cousins ever had	Banner Population			Total
	Metropolitan	*Rural*	*Floating*	
0	6,161	35,779	8,353	50,293
	(0.94)	(0.68)	(0.82)	(0.73)
1	176	6,024	847	7,047
	(0.03)	(0.12)	(0.08)	(0.10)
2	103	4,215	468	4,786
	(0.02)	(0.08)	(0.05)	(0.07)
3	38	2,619	239	2,896
	(0.01)	(0.05)	(0.02)	(0.04)

DRAFT

	Metropolitan	*Rural*	*Floating*	Total
0	6,235	45,606	9,894	61,735
	(0.96)	(0.87)	(0.97)	(0.89)
1	156	4,053	205	4,414
	(0.02)	(0.08)	(0.02)	(0.06)
2	79	1,584	49	1,712
	(0.01)	(0.03)	(0.00)	(0.02)
3	31	619	20	670
	(0.00)	(0.01)	(0.00)	(0.01)
4	11	259	2	272
	(0.00)	(0.00)	(0.00)	(0.00)
5	5	119	0	124
	(0.00)	(0.00)	0.00	(0.00)
>=6	6	105	0	111
	(0.00)	(0.00)	0.00	(0.00)
	6,523	52,345	10,170	69,038

Source: CMGPD-SC, 1866-1913.
Note: Restricted to males and unmarried daughters;

UNCLE_COUNT

brothers living in the household in the current register. For married and widowed women, UNCLE_COUNT is the number of husband's father's brothers living in the household in the current register. The values were generated by software based on the presence of men in the household whose FATHER_ID was the same as the index individual's GRANDFATHER_ID. Figure 22 shows the average number of uncles of males and unmarried daughters living in the same household by age, while Table 24 shows the distribution of the number of uncles males and unmarried daughters ever registered in the same household.

Figure 22 Average UNCLE_COUNT by Age

Source: CMGPD-SC, 1866-1913.
Note: Restricted to males with valid age (in *sui*) and PRESENT=1.

Table 24 Number of UNCLE_COUNT

#Uncles ever had	Banner Population			
	Metropolitan	*Rural*	*Floating*	Total
0	4,934	24,268	7,422	36,624
	(0.76)	(0.46)	(0.73)	(0.53)
1	978	12,543	1,298	14,819
	(0.15)	(0.24)	(0.13)	(0.21)
2	331	8,376	830	9,537
	(0.05)	(0.16)	(0.08)	(0.14)
3	153	4,384	438	4,975
	(0.02)	(0.08)	(0.04)	(0.07)
4	76	1,665	136	1,877
	(0.01)			

father's sisters living in the household in the current register. However, because almost all women have married by their twentieth and moved to a new household, the is extremely rare.

The values were generated by software based on the presence of unmarried daughters in the household whose FATHER_ID was the same as the index individual's GRANDFATHER_ID. Figure 23 shows the average number of aunts of males and unmarried daughters living in the same household by age, while Table 25 shows the distribution of the number of aunts males and unmarried daughters ever registered in the same household.

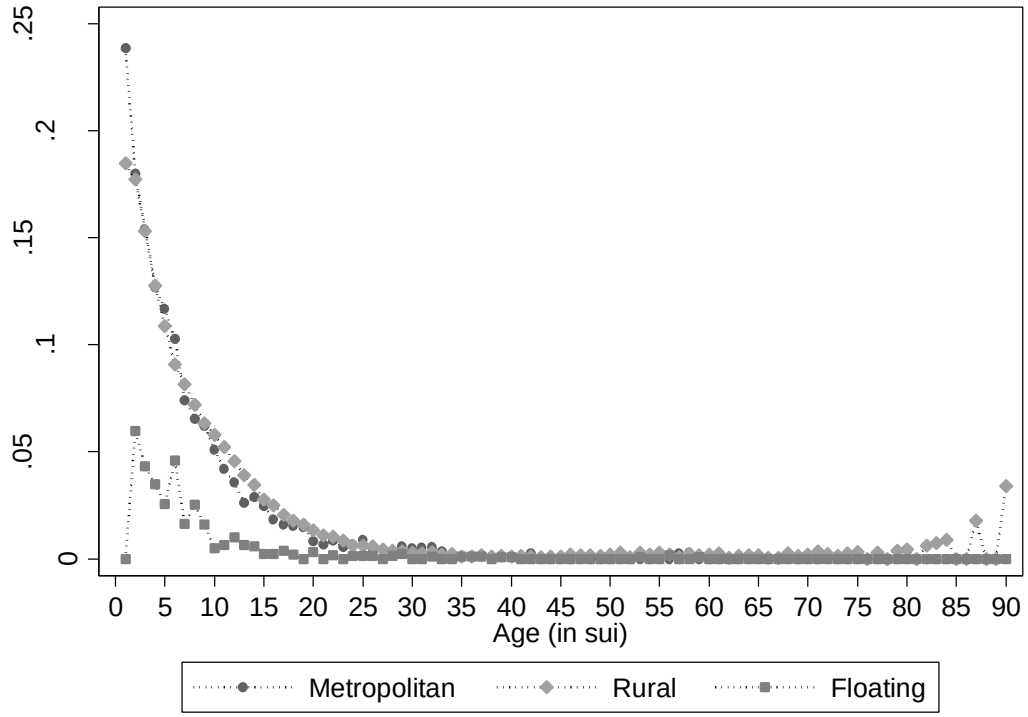


Figure 23 Average AUNT_COUNT by Age

Source: CMGPD-SC, 1866-1913.
Note: Restricted to males with valid age (in *sui*) and PRESENT=1.

Table 25 Number of AUNT_COUNT

#Aunts ever had	Banner Population			
	Metropolitan	*Rural*	*Floating*	Total
0	5,989	48,601	10,111	64,701
	(0.92)	(0.93)	(0.99)	(0.94)
1	312	2,856	52	3,220
	(0.05)	(0.05)	(0.01)	(0.05)
2	150	690	6	846
	(0.02)	(0.01)	(0.00)	(0.01)
3	43	159	1	203
	(0.01)	(0.00)	(0.00)	(0.00)
4	13	37	0	50

		(0.00)	(0.00)	0.00	(0.00)
	5	14	2	0	16
		(0.00)	0.00	0.00	(0.00)
	>=6	2	0	0	2
		(0.00)	0.00	0.00	0.00
Total		6,523	52,345	10,170	69,038

Source: CMGPD-SC, 1866-1913.
Note: Restricted to males and unmarried daughters;

3.C.II Constructs for Event History Analysis

AT_RISK_DIE

AT_RISK_DIE is a flag variable identifying observations to include in mortality analysis. The values are generated based on PRESENT and NEXT_1 (or NEXT_3). It is set to 1 if PRESENT and NEXT_1 (or NEXT_3) are both 1, which means that the observed individual is present in the current register and an observation is available in the dataset for the next annual or triennial register. AT_RISK_DIE is set to missing (-98) for individuals annotated as having exited since the last register. Otherwise the value is set to 0.

AT_RISK_MARRY

AT_RISK_MARRY is a flag variable identifying observations to include in analysis of first marriage. The values are generated by a computer program, based on the values of PRESENT, NEXT_1 (or NEXT_3), and MARITAL_STATUS. The value is set to 1 if PRESENT and NEXT_1 (or NEXT_3) are both 1 and MARITAL_STATUS is unmarried (2), which means the observed individual is present and unmarried in the current register, and an observation is available in the dataset for the next annual or triennial register. AT_RISK_MARRY is set to missing (-98) for individuals annotated as having exited since the last register. Otherwise the value is set to 0.

AT_RISK_REMARRY

AT_RISK_REMARRY is a flag variable that identifies observations to include in analysis of remarriage. The values are generated using a computer program, based on the information of PRESENT, NEXT_1 (or NEXT_3), and MARITAL_STATUS. The value of AT_RISK_REMARRY is set to 1 if PRESENT and NEXT_1 (or NEXT_3) are both 1 and MARITAL_STATUS is widowed, which means the observed individual is present and widowed in the current register, and an observation is available for them in in the next annual or triennial register. AT_RISK_REMARRY is set to missing (-98) for individuals annotated as having exited since the last register. Otherwise the value is set to 0.

NEXT_1

NEXT_1 is a flag variable. The value 1 indicates that the next annual observation of

children listed in the current and next registers because it may include births of individuals who were not recorded in the next register, but first appeared in a later register. This variable is likely to underestimate the actual number of boys because, as discussed elsewhere, many boys who died in infancy or early childhood were omitted from the registers. Accordingly, this variable may be best thought of as a count of the number of boys born between the current and next register who survived long enough to be listed in a register.

NEXT_GIRLS

NEXT_GIRLS is a count of the number of girls born to the observed individual between the current and next available registers. Please see the discussion for NEXT_BOYS on how the variable was created. Because many daughters were omitted from the original data, this variable underestimates the actual number of daughters, and should be used with extreme caution. There are pronounced patterns by time, place, and dataset in the recording of daughters. Users should familiarize themselves with these patterns by examination of relevant descriptive statistics before using this variable in an analysis.

3.C.III Family and Household

BIRTH_ORDER

This variable specifies the individual s birth order, based on comparison of calculated birth years of sons and daughters recorded as born to the same father, as indicated by the value of FATHER_ID. It is set to missing for individuals who could not be linked to a father, and for whom FATHER_ID was missing. For siblings born in the same year, ties are broken randomly. By definition, this is based on births that survived long enough to be registered, and doesn t include children who were born but died before their parents had an opportunity to register them. Thus, BIRTH_ORDER may underestimate the birth order that would be calculated if all births were recorded. Some individuals who according to BIRTH_ORDER were firstborn (i.e. BIRTH_ORDER = 1) may actually have been second or later births, whose older siblings all died before registration.

In principle, information in RELATIONSHIP should also provide birth order as it was

indicate that an individual was the head s first son s second son. That information is not used in the construction of this variable.

BIRTH_ORDER_SEX

This is the same as BIRTH_ORDER, except calculated for siblings of the same sex.

HH_SIZE

This is a count of the number of live individuals present in the household in the current register. It is based on the number of records of individuals with the same values of

24 below shows that parents tend to give diminutive names to their sons at a younger age and then switch to more dignified names when they reached adulthood. However, about 9 percent of the males still kept their diminutive names into adulthood.

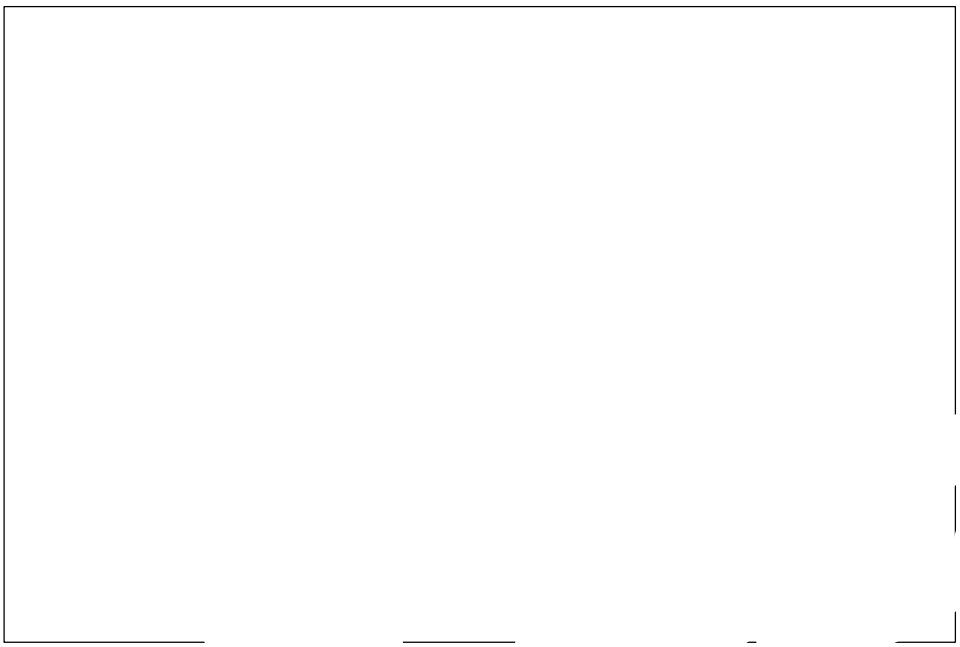


Figure 24 Proportion of males with a DIMINUTIVE given name, by age

HAS_SURNAME

This is a flag variable that indicates whether or not a male has a surname as part of the name recorded in the original register. Technically, this is based on whether there is a blank space in the pinyin for the individual s name as entered in the variable NAME in the Restricted File. Coders transcribing names from the original Chinese characters in the registers into pinyin were instructed to include a space between an individual s surname, if one was recorded, and their given name. Given names were to be transcribed into pinyin with no spaces between the characters, thus a space should only be present if there was a surname. As figure 25 shows, there was a pronounced age pattern. Boys were least likely to have a surname recorded. As men aged, they were more likely to have a surname recorded. Of the men who survived to their late seventies, more than 60 percent had a surname recorded.

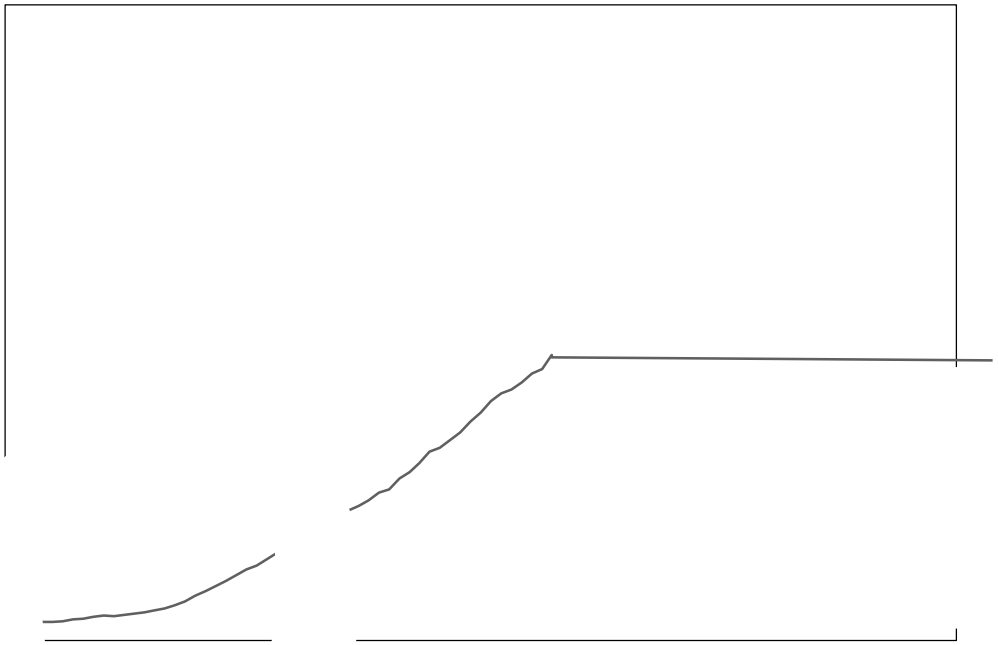


Figure 25 Proportion of Males with a Recorded Surname, by Age

NON_HAN_NAME

This variable indicates whether the given name recorded for a male was non-Han. Readers are advised to refer to the CMGPD-LN User Guide for the detailed procedure of creating this variable (Lee, Campbell and Chen 2010, p.76). In the CMGPD-SC data, 43,289 (5.45%) male observations had a non-han name. As figure 26 shows, from 1866 to 1906, the proportion of males with non-Han name declined over time. However, after 1906, this proportion slightly increased.

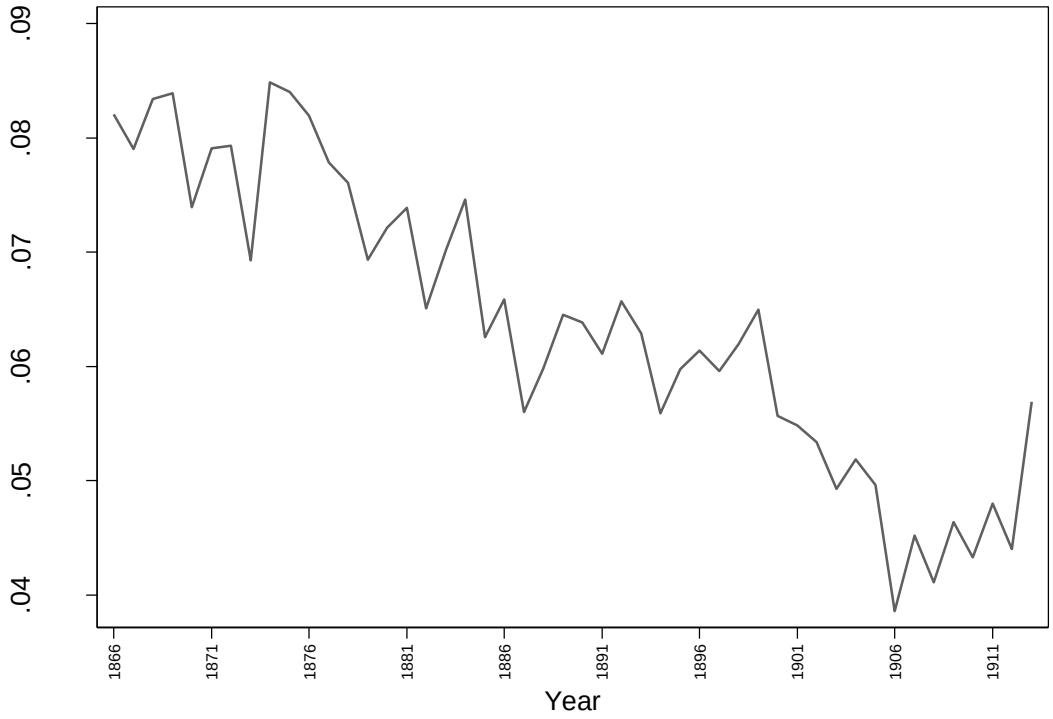


Figure 26 Proportion of males with a Non-Han given name, by year

NUMBER_NAME

Because naming children with actually numbers is a common practice among Manchu people, in CMGPD-SC, some males have given names that were actually numbers. NUMBER_NAME is a numeric variable whose contents represent the number corresponding to a male s given name if it was a number. We have included this variable to allow for investigation of the rules that parents or individuals followed in choosing the value for a number name. In CMGPD-SC, 1.43 percent of male observations had a number name.

RUSTIC_NAME

This is a flag variable indicating that a male s given name was especially suggestive of low status because they included the names of animals or medical conditions. The definition is much narrower than for DIMINUTIVE_NAME, which was set to 1 if the given name included *xiao* or *zi*. RUSTIC_NAME is set to 1 if the given name includes the pinyin for the character *gou* (dog), *zhuzi* (little pig), *niuzi* (little cow), *luzi* (little donkey), *gui* (ghost), *huazi* (pockmarked), *shazi* (fool), *xiazi* (blind), *tuzi* (bunny), or *tiaozi*, *touzi*, *tanzi* (paralyzed). We added some more names manually that did not fit these criteria, but were clearly suggestive of low status. In total, 384 of 26,641 unique male names were classified as rustic according to these criteria. After initial experiments -LN used the more broadly defined DIMINUTIVE_NAME, which yields more observations and seems to be more clearly related to social status. In CMGPD-SC, 0.76 percent of male observations had a rustic name.

3.D Identifier Variables

All the identifier variables are generated by software. In terms of their function, there are two kinds of identifier variables: those created for data management and grouping purposes; and those created for data linkage. The variables RECORD_NUMBER, REGISTER_SEQ, ZU_SEQ, and HOUSEHOLD_SEQ belong to the first category, and the variables PERSON_ID, MOTHER_ID, FATHER_ID, FATHER_ID_IMPUTED, GRANDFATHER_ID, GRANDFATHER_ID_IMPUTED, WIFE_1_ID, WIFE_2_ID, and HUSBAND_ID belong to the second category. The linkage variables are actually special features of the CMGPD_SC. Largely drawn from the links across registers made by the coders and links between kin suggested by the RELATIONSHIP variable, these variables link observations of the same individual in different registers (PERSON_ID) and link individuals to family members (MOTHER_ID, etc.)

It is important to keep in mind that the identifier variables for linkage were all generated by complex software that processed RELATIONSHIP and links made by coders, and may not be perfect. While most of the linkage is fairly straightforward, errors in the original relationship data or in the coders' transcriptions of it could lead to incorrect linkages among kin. For cases where a direct link to a kin cannot be made, the software takes an indirect approach. For example, if no father can be identified for an individual, most likely because the father passed away before that individual first appears in a register, the software checks to see if any older siblings have a father identified, and copies over that information if it is available. It may be that in some cases, that is an inappropriate assumption. All of the information that the software uses to make links between kin is available in the Basic Release in the form of the variables such as RELATIONSHIP, MARITAL_STATUS, AGE_IN_SUI, and HOUSEHOLD_ID, so users with the appropriate skills who would like to write their own software for kinship linkage are able to do so.

3.D.I Grouping Identifiers

RECORD_NUMBER

RECORD_NUMBER is a sequential record identifier identifying the location of the record within the CMGPD-SC data in its original order, when it is sorted by DATASET, YEAR and REGISTER_SEQ. Each record number identifies a unique observation in the entire dataset.

REGISTER_SEQ

REGISTER_SEQ is the sequential identifier for the records in the register in a dataset for a particular year. The value was assigned by transcribers based on interpretation of original data. Concatenating with DATASET and YEAR, it also uniquely identifies records. Sorting on DATASET, YEAR and REGISTER_SEQ restores the dataset to

for whom a surname was recorded, but there were a few cases of non-uniformity that were addressed by assigning the modal surname.

PERSON_ID

PERSON_ID is a unique identifier for individuals. Each distinct value identifies all the records of an individual in the dataset. This variable was constructed in two stages. In the first stage, the coders manually linked an individual in one register to their observation in the previous available register by specifying that observation's record number as a link identifier. Households and their members were recorded in the same order in successive registers, thus it was easy for the coders to identify and link an individual in different registers according to his/her name and contextual information of household and residential village. In the second stage, software concatenated these links from each register to the one previous to create an identifier for all of the records of an individual. There are 107,551 unique PERSON_IDs.

PERSON_ID allows users to group CMGPD-SC records by individual. Combined with YEAR, is easy for users to arrange the records in CMGPD-SC by individual and within individual by year, and copy information from one record to another. For example, `bysort PERSON_ID (YEAR): generate birthyear_first = BIRTHYEAR[1]` would create a new variable that for all the linked observations of a person would contain their calculated year of birth from the first record in which they were observed, which in turn could be used as the basis for a calculated age. Use of PERSON_ID also allows for selection of records of individuals according to specified criteria. `bysort PERSON_ID (YEAR): keep if AGE_IN_SUI[1] >= 1 & AGE_IN_SUI[1] <= 10` would throw out all records except those of people who were first observed in the registers between the ages of 1 and 10.

Because PERSON_ID is created from manually assigned links between records in adjacent registers, there are situations where records for different individuals have been linked together by mistake. There are examples of groups of records that have the same PERSON_ID but discrepancies suggestive of incorrect linkage, for example, different sex, or inconsistent ages or relationships across different records.
In a small number of situations where two or records in one register were linked to the same record in a previous register, the latter register will contain more than one record with the same value of PERSON_ID. Consistency checking by the software that produced the extract identified most such cases in the original data, and we corrected them before the release. We continue to correct such problems as we discover them or they are reported to us.

While such problems are rare enough that they should not affect analysis, users who are concerned may adjust for them by filtering records to remove inconsistencies introduced by incorrect links. For example, `bysort PERSON_ID (YEAR): keep if SEX == SEX[1]` would throw out all records where the recorded SEX was different from the one specified in the first record in the group. Similarly, `bysort PERSON_ID YEAR: keep if _n == 1` would retain only one record per PERSON_ID per YEAR. Additional restrictions might be applied in other situations where excluding other very specific types of inconsistencies is important.

Some of the situations where two or more records in one register are linked to the same record in a preceding register reflect cases in the original data where two or more records in one register clearly referred to a single individual in a preceding register. Sometimes this was the deliberate result of an adoptee being recorded in both their natal household and their adoptive household. In other cases it seems to be clerical error.

There are also examples of situations where an improbably large number of records have the same value of PERSON_ID because a person s death was never recorded in the register, usually because they **Error! Reference source not found.**, and they ere carried forward from one register to the next indefinitely. See the discussions of **Error! Reference source not found.** and **Error! Reference source not found.** for re discussion of this phenomenon.

When such inconsistencies across registers reflected peculiarities in the original data, we left everything as is in the release, rather than try to fix the contents of the records or change the links. Thus even if we release versions of the data that correct problems that are the result of transcription mistakes by the coders, future releases will not correct problems that are in the original sources.

The implications of these issues will depend on what the user seeks to do with the data. Problems with PERSON_ID and inconsistencies in records that share the same value of PERSON_ID are rare enough that for most applications, the data may be used as is.

More serious issues may arise when data management involving complex merge and sort that assume that in a given year, each value of PERSON_ID appears only once. If this assumption is violated, it may lead to odd behavior, such as failed `merge` operations, unexpected appearance of new records following a `merge`, or variables created through `bysort` not having expected values. In general, the best approach to these more complex issues is to eliminate the typically small number of badly behaved records before carrying out more complex operations that make strong assumptions about the uniqueness of combinations PERSON_ID and YEAR. For example, the following would eliminate all individuals as defined by a common value of PERSON_ID who had two or more records in any one year:

```
bysort PERSON_ID YEAR: generate duplicates = _N
     bysort PERSON_ID (duplicates): drop if
duplicates[_N] > 1
```

The distribution of linked observations per individual recorded is presented in Table 26. The peculiar distributional patterns for male and female floating bannermen are driven by the fact the floating banner population was no longer registered in later years. There are a total of 82,535 observations of individuals who were recorded in 36 or more registers. Also note that 647 individuals are recorded to have duplicated observations for a single year and are excluded from the calculation.

Table 26 Distribution of linked observations per individual recorded

	Metropolitan	*Rural*	*Floating*	Total

1-5	7,993	66,705	18,632	93,330
	(0.06)	(0.06)	(0.15)	(0.07)
6-10	11,996	131,783	69,671	213,450
	(0.09)	(0.12)	(0.54)	(0.16)
11-15	17,872	160,783	39,751	218,406
	(0.13)	(0.15)	(0.31)	(0.16)
16-20	21,824	158,632	312	180,768
	(0.16)	(0.15)	(0.00)	(0.13)
21-25	16,280	194,120	104	210,504
	(0.12)	(0.18)	(0.00)	(0.16)
26-30	11,132	193,746	10	204,888
	(0.08)	(0.18)	(0.00)	(0.15)
31-35	13,246	129,680	19	142,945
	(0.10)	(0.12)	(0.00)	(0.11)
36+	37,407	45,128	0	82,535
	(0.27)	(0.04)	0.00	(0.06)
Total	137,750	1,080,577	128,499	1,346,826

Source: CMGPD-SC, 1866-1913.
Note: Individuals with duplicated annual records are excluded;
[2] A very small number of individuals who are missing on sex are also excluded.

MOTHER_ID

MOTHER_ID provides the PERSON_ID of mother for males and never-married females or the PERSON_ID of the mother-in-law for married or widowed women. This variable is generated by software that first seeks to link individuals to mothers based on relationships recorded in the household registers. For example, if the software finds a `1s1s`, it will look for a `1sw` in the household. The software carries out the searches for the same individual in all the registers in which they appear and in the case of apparent conflicts between registers, gives priority to the link made in the earliest register. Once the software exhausts possibilities for direct linkage based on relationship, it seeks indirect links, for example, by checking whether an individual identified as a father had a wife who died before the individual first appeared in the registers, or whether a sibling had a mother identified. The value for individuals for whom the software was unable to locate a mother is set to -99.

FATHER_ID

FATHER_ID provides the PERSON_ID for the father of males and never-married

Thus, for married or widowed women, the person identified by FATHER_ID is their father-in-law. The values of FATHER_ID are generated by a computer program in two ways. For individuals whose father is in the original data, the values are the real PERSON_ID of father, assigned by linking father to an individual based on their relationship recorded in the household registers. For individuals whose father is not in the original data, the value is imputed.
There are 2,426 (2.25%) individuals who do not have a valid FATHER_ID and are

thus coded as -99.

FATHER_ID_IMPUTED

FATHER_ID_IMPUTED is a flag variable that indicates that the FATHER_ID in this observation refers to an individual not in the dataset. FATHER_ID values were generated to group individuals who were clearly siblings based on the values of RELATIONSHIP, but who could not be linked to a father in the dataset, generally because they were in very early registers and their father had already died without ever being listed in a register included in the CMGPD-SC. For example, if a group of siblings was observed living together in an earliest available register, without a father, a common father was assumed for them and an identifier assigned to him that would allow his offspring to be grouped together during analysis, even though he did not appear anywhere in the registers. If FATHER_ID_IMPUTED is 1, FATHER_ID may be used to group observations of children of the same father, but may not be used to link to the father's observations to gather information about him. FATHER_ID is imputed for 23.04% of individuals with non-missing FATHER_ID.

GRANDFATHER_ID

GRANDFATHER_ID provides the PEwNON_ID of paternal grandfather for males and never-
or widowed women. Like FATHER_ID, the values of GRANDFATHER_ID are also generated by software and indicate paternal grandfather-in-law for married or widowed women. For individuals whose grandfathers can be identified in the raw data and can be located by a search on RELATIONSHIP, the values are the original PEwNON_ID of the grandfather turned up by that search. For example, a grandson of a head, `1s1s`, can be linked directly to his grandfather, the head, `e`. Where a grandfather could not be identified directly from the data, most likely because he had already died, grandfather was assumed to be the individual's father's father, so FATHER_ID from the father was copied to GRANDFATHER_ID for the individual.

GRANDFATHER_ID_IMPUTED

GRANDFATHER_ID_IMPUTED is a flag variable that indicates the GRANDFATHER_ID in this observation refers to an individual not recorded in the original data, but whose existence was inferred and for whom an identifier was assigned to allow for grouping of grandchildren. See FATHER_ID_IMPUTED for an explanation. If GRANDFATHER_ID_IMPUTED is set to 1, GRANDFATHER_ID may only be used to group observations of individuals who had a common paternal grandfather.

For 43,753 (40.5%) out of the 108,020 individuals, GRANDFATHER_ID refers to an individual not recorded in the original dataset as the flag GRANDFATHER_ID_IMPUTED indicates. Of 81,043 individuals who can be linked to their fathers (FATHER_ID_IMPUTED=0), 58.7% can be linked to their grandfathers (GRANDFATHER_ID_IMPUTED=0).

WIFE_1_ID

The software generates WIFE_1_ID by linking wives to their husband based on the RELATIONSHIP recorded in the original data in the current register. For every female with a `w` at the end of RELATIONSHIP, the processing software searched the household for a man with the same RELATIONSHIP, but without a `w` at the end. For example, if a woman has RELATIONSHIP `2ybw`, second younger brother's wife, the program searched the household for a man with RELATIONSHIP `2yb`. For a head's wife, RELATIONSHIP `w`, the software searched for an `e`. For a head's mother, `m`, the software searches for a father, `f`. When the software found a match, it copied the woman's PERSON_ID into the male's WIFE_1_ID, and copied the male's PERSON_ID into the woman's HUSBAND_ID.

Because of the possibility of polygyny and widower remarriage, the software does not attempt to adjudicate between different wives identified for the same individual in different or even the same registers. If two wives are associated with an individual in the same register, the PERSON_ID of the second will be moved to WIFE_2_ID. If a different wife is associated with an individual in a later register, WIFE_1_ID in that register will be different from WIFE_1_ID in the current register.

Because WIFE_1_ID and WIFE_2_ID are filled in based on the contents of the current register, users seeking to collect data on a deceased wife's characteristics for a widower will need to carry out additional processing to copy the WIFE_1_ID for the widower forward from the register in which the wife was listed recorded, and merge based on that value for WIFE_1_ID.

WIFE_1_ID and WIFE_2_ID are set to missing (-98) for women, widowers and unmarried men.

WIFE_2_ID

WIF
raw data. Like WIFE_1_ID, the values are generated with computer programs by linking a wife to a husband based on the relationship recorded in the raw data. Basically, if more than one woman in a household had a RELATIONSHIP that matched the same male, the PERSON_ID of the second match was copied over to WIFE_2_ID for the male. Polygyny was extremely rare in the populations covered in the registers, thus there are very few such cases.

HUSBAND_ID

register. It is generated as part of the processing based on RELATIONSHIP that identifies WIFE_1_ID and WIFE_2_ID. Once again, it only refers to the husband identified through RELATIONSHIP in a current register.

HUSBAND_ID is set to missing (-98) for women with MARITAL_STATUS of

widowed or unmarried. It is also set to missing (-98) for all men. It is set to missing (-99) for women whose MARITAL_STATUS was married, but for whom a husband could not be located in the current register. Most of these were women whose MARITAL_STATUS was married because they were daughters annotated as having married out since the last register.

KIN GROUP VARIABLES

The kin group variables are intended for grouping observations of individuals who are related to each other through a common patrilineal ancestor. These variables allow for patrilineal kin groups, variously defined, to be treated as units of analysis. We provide several variables that apply different definitions for the kin group, ranging from the more narrowly defined to the more broadly defined. The most narrowly defined kin group is the one defined by values of FOUNDER_ID, followed by FOUNDER_INFERRED_ID, then UNIQUE_YI_HU and finally UNIQUE_GROUP. UNIQUE_HH_ID is constructed

different registers, and therefore is not directly comparable to the other kin group variables described in this section.

FOUNDER_ID

FOUNDER_ID is assigned based on the earliest patrilineal ancestor located for an individual in the registers. Men who are all traced back to the same patrilineal ancestor observed in the earliest available register in a DATASET will have the same value of FOUNDER_ID. Combined with YEAR, FOUNDER_ID can be used to group observations of individuals in each year who share the same patrilineal ancestor. Wives and widows are assigned the same value of FOUNDER_ID as their husbands. FOUNDER_ID was constructed for each individual by linking back from one generation to the previous by chaining together values of FATHER_ID until a male was located who represented the earliest person actually observed in the registers. The associated programming was quite complex, and involved many decisions about handling contradictory or unclear situations, thus it is possible that someone working independently writing programs to achieve the same goal would yield slightly different results, depending on how they handled special cases. There are 16,727 distinct values of FOUNDER_ID, corresponding to 16,866 distinct descent lines defined by common descent from a patrilineal ancestor in the earliest available register.

FOUNDER_INFERRED_ID

Each unique value of FOUNDER_INFERRED_ID groups together observations of men who can all be traced to a common patrilineal ancestor whose existence can be inferred from processing the values of RELATIONSHIP for men in the same household. In principle, it should define a broader kin group than FOUNDER_ID, which is based on descent from a patrilineal ancestor observed in the register.

the men s relationships to each other. For each of the two brothers, their descendants will have different values of FOUNDER_ID according to which brother they were descended from, since FOUNDER_ID identifies common patrilineal descent from an individual recorded in the registers. Descendants of men who were identified as cousins, second cousins, or other patrilineal kin by RELATIONSHIP in the earliest available register will all have the same value of FOUNDER_INFERRED_ID. Again, wives inherit values of FOUNDER_INFERRED_ID from their husbands.

Because this variable relies heavily on a complex algorithm to process the strings in RELATIONSHIP to identify individuals who may be related in the sense of having a common patrilineal ancestor, it should be treated as experimental, and used with caution. There are actually 30,896 distinct values of FOUNDER_INFERRED_ID, which almost doubles the number for FOUNDER_ID. However, the median number of observations for groups defined by distinct values of FOUNDER_INFERRED_ID is 102, whereas the median number of observations defined by distinct values of FOUNDER_ID is 200.

UNIQUE_GROUP

Values of UNIQUE_GROUP represent the broadest definition of kin groups. For the purposes of this calculation, members of *yihu* with the same surname listed consecutively in an earliest available register in a DATASET were assumed to be descended from a common patrilineal ancestor. This assumption is based largely on the observation that men in families in adjacent *yihu* who have the same surname also tend to share the same generational characters. There are 4,124 distinct values of UNIQUE_GROUP. Thus, according to the definition used to construct this variable, the CMGPD-SC individuals can be divided into 4,124 distinct kin groups based on common descent from a patrilineal ancestor. These patrilineal ancestors were not recorded in the extant registers but may have lived quite some time before the earliest available register. It remains possible that some kin groups that have different values of UNIQUE_GROUP actually share common descent from an even earlier male ancestor, but this common descent could not be discerned by the procedures for automated linkage that generated values of UNIQUE_GROUP.

UNIQUE_HH_ID

UNIQUE_HH_ID is intended to group observations of individuals who lived in the same household recorded in consecutive registers. Since households evolve over time as the result of the entrance and exit of individual members, as well as processes of household division, this required imposition of a definition of household that could be interpreted in a longitudinal context. For the purposes of creating this variable and assigning values, we chose to define a household in a longitudinal context as one in which individuals may have entered or exited through death, birth, marriage, or other processes, but there were no household divisions (individuals who were in the same household in one register appeared in separate households in the following register). When a household divided, we s and assigned each their own value of UNIQUE_HH_ID, which they retained until they divided again. At least in principle, individuals represented in observations with the same value of UNIQUE_HH_ID but in different years should have shared a similar household environment, save for the entry and exit of individual household members, and other unobserved changes in the household context. If the household head changes from one register to the next,

UNIQUE_VILLAGE_ID

Each village is uniquely identified by its UNIQUE_VILLAGE_ID, which corresponds to a unique combination of the address of the village and banner affiliation recorded in the household registers with the following adjustments:
(1) Plain Yellow metropolitan bannermen (DATASET=101) have been combined with their Plain Red rural bannermen (DATASET=105) counterparts with the same village address;
(2) Bordered Yellow metropolitan bannermen (DATASET=102) and Plain White rural bannermen (DATASET=103) with the same village address are also combined;
(3) rural bannermen and floating bannermen sharing the same banner and village address are combined unless the address is a *wopeng* ( ), or cottage which is coded as 0. There are 139 unique values of UNIQUE_VILLAGE_ID.

Each village is assigned a unique UNIQUE_VILLAGE_ID, including 9 village IDs created in an *ad hoc* way because of missing information on village address (See Appendix Table A). All such *ad hoc* village IDs with a relatively large number of observations are for the floating banner population, which was not organized by village. Therefore, will not pose a problem if the whole floating banner group is excluded from analyses by village.

UNIQUE_VILLAGE_NUMBER

The UNIQUE_VILLAGE_NUMBER is created the same way as the UNIQUE_VILLAGE_ID. The major difference between the two variables is that the values of UNIQUE_VILLAGE_NUMBER keeps the information of the banner (DATASET) and village address, while the values of UNIQUE_VILLAGE_ID rang from 1 to 139.

LATITUDE

LONGITUDE

ORIGINAL_ADMIN

ORIGINAL_ADMIN is the banner organization to which the immigrant household belonged before moving to SC.

ORIGINAL_COMMANDER

ORIGINAL_COMMANDER is the place name
banner organization was located before moving to SC. Each value of the variable represent a place name, which is available upon request.

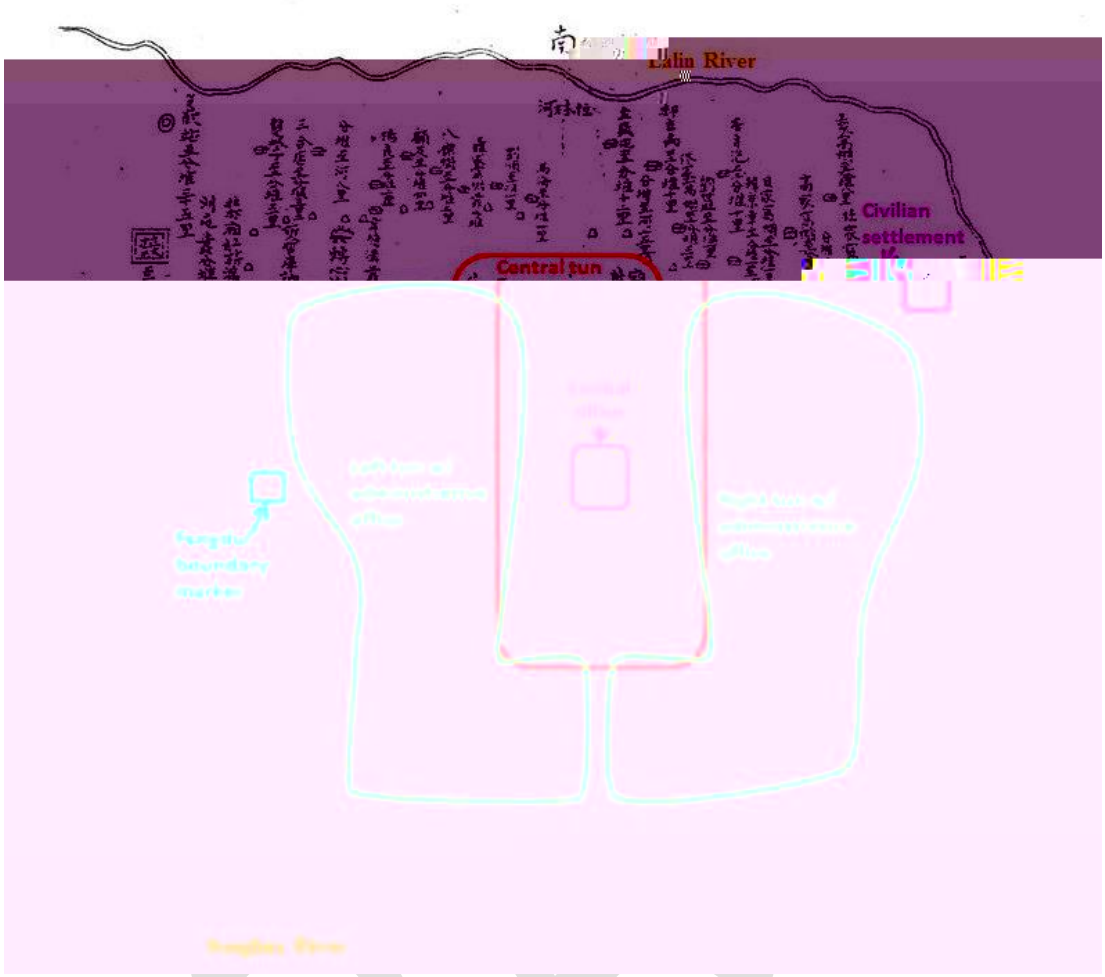
NEW_ADDRESS

NEW_ADDRESS is the village address to which the individual and his direct family members moved after receiving a new plot of land.

example, a captain, who usually supervised one banner elsewhere, administered four banners in Shuangcheng.

In the end of 1869, the state reorganized the banner units in Shuangcheng and consolidated the three sets of eight banners into one. The government first introduced one intermediate unit in the Eight Banners *jiala* to replace banner as the name of the five-village unit. In a related move, the 20 villages administered by one captain, which originally comprised four banners, became only one banner. With this reorganization, although the village boundaries and the unit of register compilation the 20 villages under one captain remained the same, the names of the villages changed. A village was named after the *jiala*
jiala, for example, the fifth village of the third *jiala* of the Bordered Yellow banner.

Beginning in 1870, the eight banners in Shuangcheng were no longer a mere reflection of geographical boundaries, but also an indicator of the category of the population it administered. In terms of population category, all the metropolitan banner households were organized under the two yellow banners: plain yellow and bordered yellow, and all the rural and floating banner households were organized under the other six banners: plain white, bordered white, plain red, bordered red, plain blue, and bordered blue. In terms of geographical boundaries, each banner administered 20 villages. Therefore, four banners plain yellow, bordered yellow, plain white, and plain red administered the 40 villages in the central *tun;* two banners bordered red and bordered blue administered the forty villages in the right *tun*; and two banners bordered white and plain blue administered the forty villages in the left *tun.* The banner administration also separated floating bannermen from rural bannermen when compiling population registers.

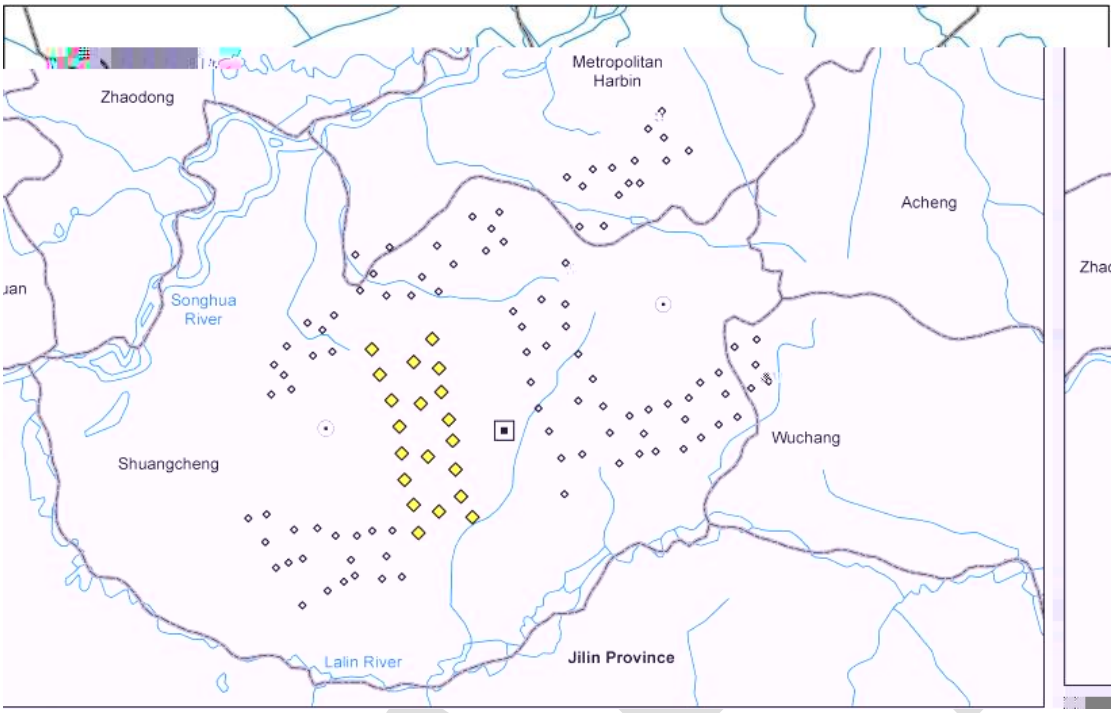


Map 4 Organization of Shuangcheng Banner Villages
Source: SCPTTJL, back cover.

Thereafter, each year there were 14 distinct register books for the Shuangcheng banner households, identifying 14 distinct populations associated with population status and geographical boundaries. In this section, we summarize the characteristics of each of these 14 banner populations in the CMGPD-SC (See Table).

4.A Plain Yellow Banner Metropolitan Bannermen (*zhenghuangqi jingqi*) (DATASET 101)

The Plain Yellow Banner in Shuangcheng administered the metropolitan banner households living in the twenty villages west of Shuangcheng City (Map 5). The population registers were organized first by village and then by household under the overall administration of a captain (*zuoling*) and four chief village heads called *zong tunda,* each supervising five villages. Each village also had its own village head called *tunda.* The Plain Yellow Banner had a population of 1,105 in 1866, and increased to 2,349 in 1912.



Map 5 Locations of Plain Yellow Metropolitan Banner Population

Due to the association of households with land plots allocated by the state, the number of registered households under the Plain Yellow Banner was fixed according to the number of allocated land plots. From 1866 to 1869, there were around 260 separately registered households which increased to 350 in 1870, 498 in 1879, and 500 in 1880 as the state reallocated land plots left behind by extinct metropolitan banner households in 1869 and allocated 332 plots of land to the metropolitan bannermen living in the 40 villages in the central *tun* in 1878.[26] After 1880 there were no new land allocations to this population, and the number of 500 registered households remained unchanged until 1912.

As Table 28 shows, all the metropolitan bannermen under the Plain Yellow Banner were from Beijing, and were registered as Manchu, Mongol, and Xibe.[27]

Table 28 **Characteristics of the Plain Yellow Banner metropolitan banner population**

Population category		Metropolitan (*jingqi*)	
Ethnicity		Obs.	%
	Manchu	13,697	77.33
	Xibe	194	1.1
	Mongol	3,821	21.57
	Total:	17,712	100
Organization		Organized by household *(linghu)*	

[26] Please see Chen 2009 chapters 6 and 7 for the details of the history of land allocation in Shuangcheng.
[27] The metropolitan bannermen only had a handful of households of Xibe ethnicity.

Administrative authority	Captain (*zuoling*), Chief village head (*zong tunda*), and Village head (*tunda*)
National Exam Title*	0.5 percent
Government Employment*	10.5 percent

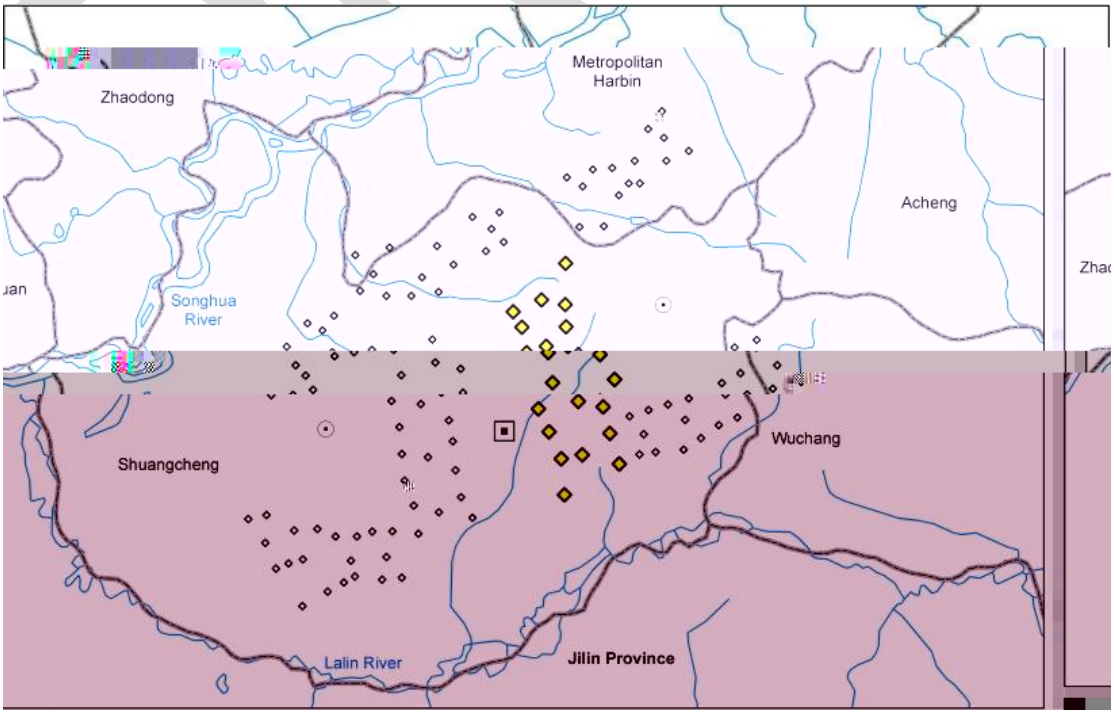
Source: CMGPD-SC, 1866-1913.
Note: *The denominator is male observations between 18 and 60 *sui*.

Compared to the males in the CMGPD-LN data, the males in the CMGPD-SC Plain Yellow Banner had a low percentage of members participating in state sponsored exams. Only 0.5 percent of the males between 18 and 60 *sui* participated in state sponsored exams. However, 10.5 percent of males of this age group held salaried positions, which is very high compared to the CMGPD-LN population and to rural and floating bannermen living in Shuangcheng.

4.B Bordered Yellow Banner metropolitan bannermen (*xianghuangqi jingqi*) (DATASET 102)

Because they belonged to the same population category of metropolitan bannermen, the population under the Bordered Yellow Banner (*xianghuang qi*) shared many features of those under the Plain Yellow Banner. These metropolitan banner households lived in the 20 villages east of the seat of Shuangcheng (Map 6). The population registers were organized first by village and then by household. A captain and four chief village heads supervised the 20 villages. The population of this banner increased from 1,029 in 1866 to 2,250 in 1912.

Like the Plain Yellow Banner population, the number of registered Bordered Yellow Banner households increased from 234 households in 1866 to 347 in 1870, 350 in 1871, 487 in 1879, and 500 in 1884 in response to two new land allocations in 1869 and in the late 1870s.



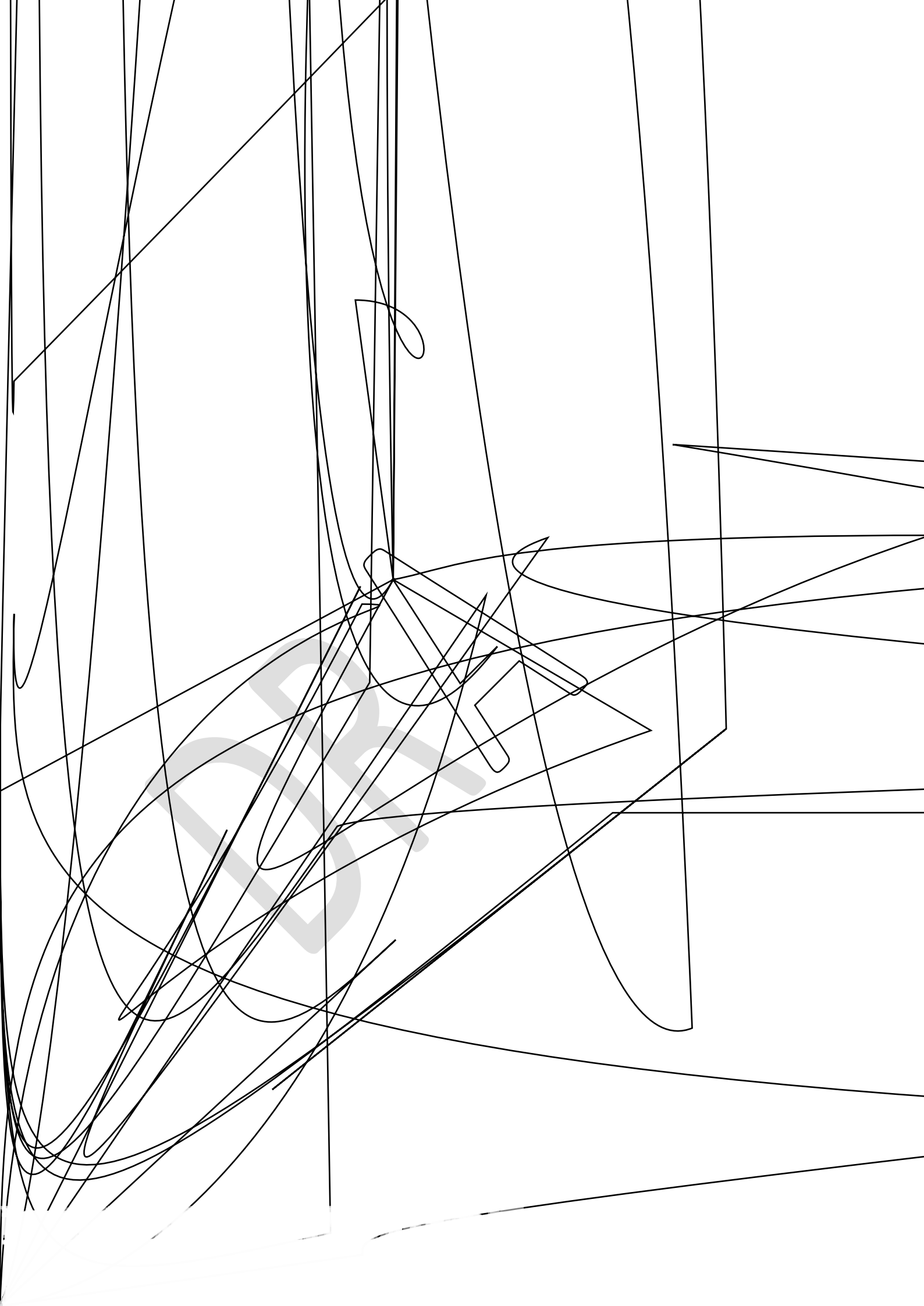
Map 7 Locations of Plain White Rural Banner Population

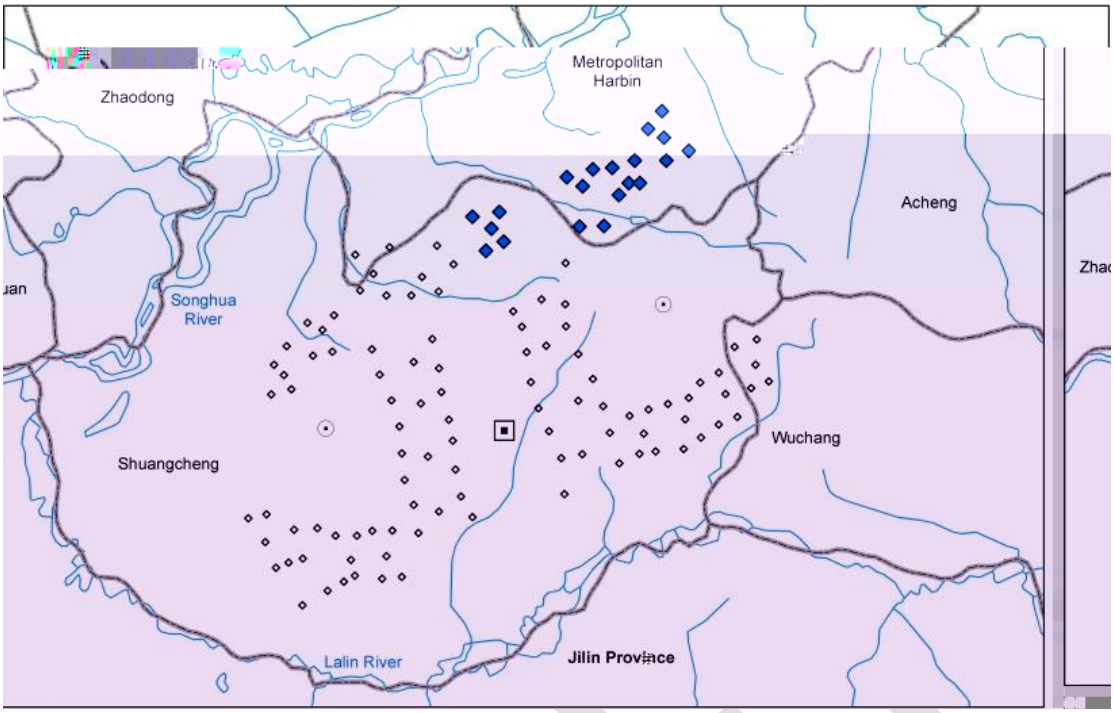
Because the government completed land allocation to rural banner households soon after their arrival in Shuangcheng in 1820, the number of households remained relatively stable throughout the time. For the period covered by the CMGPD-SC, the number of rural households of the Plain White banner was 480 in 1868 and 503 in 1869. In 1910, the number of households reached 523.[28]

All the households administered under the Plain White Banner came from a number of locations in Liaoning and Jilin provinces. In the registers, the households were cat Manchu, Mongol, Xibe, Han, and Baerhu, in order of status. Compared to other rural banner populations in the CMGPD-SC, the Plain White Banner rural population was the smallest, with 11,040 individuals in total (Table 1).

Table 30 Characteristics of the Plain White banner rural bannermen population

Population category	Rural (*tunding*)
Ethnicity	Obs.


DRAFT



Map 11 Locations of Plain Blue Rural Banner Population

The households of the Plain Blue banner consist of four ethnic groups: Manchu, Mongol, Xibe, and Han. About 0.3 percent of the males between 18 and 60 *sui* participated in state sponsored exams, and 1 percent of the males of this age group had a salaried position, which placed the Plain Blue banner the lowest among all the rural banner populations in terms of occupational attainment.

Table 34 Characteristics of the Plain Blue banner rural bannermen population

Population cateogry	Rural (*tunding*)	
Ethnicity	Obs.	%
Manchu	17,905	29.3
Han	30,899	50.56
Xibe	8,814	14.42
Mongol	2,492	4.08
Ba er hu	218	0.36
Taimanzi	785	1.28
Total:	61,113	100
Organization	Organized by household *(linghu)*	
Administrative authority	Captain (*zuoling*), Chief village head (*zong tunda*), and Village head (*tunda*)	
National Exam Title*	0.3 percent	
Government Employment*	1 percent	

Source: CMGPD-SC, 1866-1913.
Note: * The denominator is male observations between 18 and 60 sui.

4.H Bordered Blue Banner rural bannermen (*xianglanqi tunding*) (DATASET 108)

The Bordered Blue banner administered the 20 villages located in the southwest part of the state farm, in the right *tun* (Map 12). Households living in these 20 villages

Total:	64,740	100
Organization	Organized by household *(linghu*)	
Administrative authority	Captain (*zuoling*), Chief village head (*zong tunda*), and Village head (*tunda*)	
National Exam Title*	0.6 percent	
Government Employment*	2.3 percent	

Source: CMGPD-SC, 1866-1913.
Note: * The denominator is male observations between 18 and 60 sui.

4.I Plain White Banner floating bannermen (*zhengbaiqi fuding*) (DATASET 111)

The Plain White banner floating bannermen included the floating banner households living in the 20 villages east of the seat of Shuangcheng (Map 7). The captain of the Plain White banner supervised all these floating banner households. However, below the captain, there was no village head specially appointed to supervise them.[29] On the registers, the floating bannermen was not organized by village but directly by household.

Due to their lower socioeconomic and political status in Shuangcheng, the floating bannermen differed from metropolitan and rural bannermen in terms of registered population size, geographical mobility, and occupational attainment measured by salaried positions. Because floating bannermen were excluded from land allocation and were unofficial immigrants, they were allowed to emigrate from Shuangcheng, and therefore had greater geographical mobility. Moreover, because the government stopped systematically registering newly arrived floating bannermen after 1847, the registered floating banner population declined over time.

All the above factors rendered a different profile to the Plain White banner floating bannermen as well as other floating banner populations. In 1867, there were 316 households of floating bannermen under the Plain White banner. The number of households in this population significantly dropped to 202 in 1901 and further declined to 171 in 1909. The population size of the Plain White banner floating bannermen first increased from 468 in 1867 to 1,730 in 1873. Then the population gradually declined to 639 in 1901 and 581 in 1909.

Because they shared the place of origin with rural bannermen from Liaoning, the floating bannermen in the Plain White banner had a similar ethnic composition to their rural banner counterpart. This population consisted of six ethnic groups: Manchu, Mongol, Xibe, Han, Baerhu, and Taimanzi (Table 36). Almost no males between 18 and 60 *sui* in the Plain White banner floating banner population participated in state sponsored exams or held a salaried position.

[29] It is likely that the village heads supervising rural banner households also supervised the floating banner households.

Table 36 Characteristics of the Plain White banner floating bannermen population

Population category	Floating	
Ethnicity	Obs.	%
Manchu	3,017	42.6
Han	2,546	35.95
Xibo	936	13.22
Mongol	484	6.83
Ba er hu	99	1.4
Total:	7,082	
Organization	Organized by household	
Administrative authority	Captain and Household head	
National Exam Title*	0 percent	
Government Employment*	0 percent	

Source: CMGPD-SC, 1866-1913.
Note: * The denominator is male observations between 18 and 60 sui.

4.J Bordered White Banner floating bannermen (*xiangbaiqi fuding*) (DATASET 112)

The Bordered White banner floating bannermen population lived in the 20 villages located in the southeast part of the state farm, in the left *tun* (Map 8). The captain of the Bordered White banner supervised them. Similar to that of the floating bannermen in the Plain White banner, this population was directly organized by household.

The Bordered White banner floating banner population is the largest floating banner population, with 5,174 individuals over the study period. This population also underwent significant decline due to out-migration. In 1870, there were 4,123 people living in 828 households. In 1897, the population still present in Shuangcheng declined to 3,155 and the number of households in the register declined to 810.

This population also came from various places in Liaoning. It consisted of six ethnic groups: Manchu, Mongol, Xibe, Han, Baerhu, and Taimanzi (Table 37). Very few men held a salaried position, and only 0.1 percent of males between age 18-60 *sui* participated in state sponsored exams.

Table 37 Characteristics of the Bordered White floating banner population

Population category	Floating	
Ethnicity	Obs.	%
Manchu	3,097	23.84

CMGPD-

Administrative authority	Captain and Household head
National Exam Title*	0.2 percent
Government Employment*	0.1 percent

Source: CMGPD-SC, 1866-1913.
Note: * The denominator is male observations between 18 and 60 *sui*.

4.L Bordered Red banner floating bannermen (*xianghongqi fuding*) (DATASET 114)

The Bordered Red banner floating banner population lived in the 20 villages located in the northwest part of the state farm, in the right *tun* (Map 10). This population all came from Liaoning. The captain of the Bordered Red banner supervised them. They were organized by household in the registers.

Over time, the size of the Bordered Red banner declined more than 50 percent. There were 287 households of 1,548 individuals in 1867. In 1891, only 190 households of 884 individuals were registered. The number of households further declined to 129 in 1894 and 102 in 1901. The population size also declined to 689 in 1901.

The Bordered Red banner floating population consists of four ethnic groups: Manchu, Mongol, Xibe, and Han (Table 39). Only 0.1 percent of the males between ages 18 and 60 *sui* had government employment. Almost no men participated in state sponsored exams.

Table 39 Characteristics of the Bordered Red

4.M Plain Blue banner floating bannermen (*zhenglanqi fuding*) (DATASET 115)

The Plain Blue banner floating bannermen population lived in the 20 villages located in the northeast, a part of the left *tun* (Map 11). The captain of the Plain Blue banner supervised this population. These households all came from Liaoning. They were organized by household in the registers.

The size of this population also underwent significant decline. There were 619 households of 3,558 individuals in 1867. By 1909, there were only 185 households of 1,211 individuals. This population consists of six ethnic groups: Manchu, Mongol, Xibe, Han, Baerhu, and Taimanzi (Table 40). No males between age 18 and 60 *sui* participated in state sponsored exams or had government employment.

Table 40 Characteristics of the Plain Blue banner floating bannermen population

Population category	Floating	
Ethnicity	Obs.	%
Manchu	3,595	25.51
Han	8,178	58.04
Xibe	1,885	13.38
Mongol	432	3.07
Total:	14,090	
Organization	Organized by household	
Administrative authority	Captain and Household head	
National Exam Title*	0 percent	
Government Employment*	0 percent	

Source: CMGPD-SC, 1866-1913.
Note: * The denominator is male observations between 18 and 60 *sui*.

4.N Bordered Blue banner floating bannermen (*xianglanqi fuding*) (DATASET 116)

The Bordered Blue banner floating bannermen population lived in the 20 villages located in the southwest part of the state farm, in the right *tun* (Map 12). All the households came from Liaoning. The captain of the Bordered Blue banner supervised these households.

The population numbered 1,484 individuals in 286 households in 1867 and soon declined. In 1909 although the number of households only slighted dropped to 214, the number of individuals dropped to 790, a decline of about 50 percent.

This population consists of six ethnic groups: Manchu, Mongol, Xibe, Han, Baerhu, and Taimanzi (Table 41). Throughout the time, no male between 18 and 60 *sui*

participated in state sponsored exams, and only 0.2 percent of the males of this age group had government employment.

Table 41 Characteristics of the Bordered Blue floating banner population

Population category	Floating	
Ethnicity	Obs.	%
Manchu	3,034	49.71
Han	1,598	26.18
Xibe	1,284	21.04
Mongol	181	2.97
Tai man zi	7	0.11
Total:		
Organization	Organized by household	
Administrative authority	Captain and Household head	
National Exam Title*	0 percent	
Government Employment*	0.2 percent	

Source: CMGPD-SC, 1866-1913.
Note: * The denominator is male observations between 18 and 60 *sui*.

REFERENCES

Bengtsson, Tommy, and Martin Lindström. "Airborne infectious diseases during infancy and mortality in later life in southern Sweden, 1766-1894." *International Journal of Epidemiology*, 2003: 286-294.

Bengtsson, Tommy, and Martin Lindström. "Childhood misery and disease in later life: The effects on mortality in old age of hazards experienced in early life, southern Sweden, 1760-1894." *Population Studies*, 2000: 263-277.

Bengtsson, Tommy, Cameron Campbell, and James Lee. *Life Under Pressure: Mortality and Living Standards in Europe and Asia.* Cambridge: MIT Press, 2004.

Campbell, Cameron, and James Lee. "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*, 1996: 297-328.

Campbell, Cameron, and James Lee. "Economic Conditions and Male First Marriage in Northeast China, 1749-1909." *Sungkyun Journal of East Asian Studies*, 2008b: 17-42.

Campbell, Cameron, and James Lee. "Free and unfree labor in Qing China: Emigration and escape among the bannermen of northeast China, 1789-1909." *The History of the Family: An International Quarterly*, 2001: 455-76.

Campbell, Cameron, and James Lee. "Households, household groups, and individual outcomes in Liaoning, 1789-1909." In *Family Structures, Demography and Population. A Comparison of Societies in Asia and Europe*, by Catherine Capron and Muriel Neven, 111-130. Liège: Laboratoire de Démographie de

Campbell, Cameron, and James Lee. "Kinship, Employment and Marriage: The Importance of Kin Networks for Young Adult Males in Qing Liaoning." *Social Science History*, 2008a: 175-214.

Campbell, Cameron, and James Lee. "Price fluctuations, family structure and mortality in two rural Chinese populations: Household responses to economic

2610.9-6(e)4(th i)-3-6(eild ghtee d)-7(enthaon)] TJETBT1 0 0 17322.97 0Tm[(261[(-)] TJETBT1 0 0 21497.97 0Tm[(261)4

Chen, Shuang, Cameron Campbell, and James Lee. "Vulnerability and Resettlement: Mortality Differences in Northeast China by Place of Origin, 1970-1912, Comparing Urban and Rural Migrants." *Annales de démographie historique*, 2005.

Clark, Sam, Elizabeth Colson, James Lee, and Thayer Scudder. "Ten thousand Tonga: a longitudinal anthropological study of southern Zambia, 1956-1991." *Population Studies* 49, no. 1 (March 1995): 91-109.

Colson, Elizabeth. *The Social Consequences of Resettlement.* Manchester: University of Manchester Press, 1971.

Costa, Dora. "Understanding the Twentieth-Century Decline in Chronic Conditions among Older Men." *Demography*, 2000: 53-72.

Dong, Hao, and James Lee. "Kinship matters: Long-term mortality consequences of childhood migration, historical evidence from northeast China, 1792-1909." San Francisco, 2012.

Harvard Yenching Institute. *China Historical Geographic Information System (CHGIS).* Version 4. Cambridge, MA, Jan 2007.

Komekura, Niro. "Shin dai hokuman no ton kon -- Shuan cheng pu tonden no yosatsu hokoku wo chushin to shi te (Northern Manchurian frontier settlements: with a focus on a preliminary report from the Shuangchengpu state farm)." *Toa jin bun gaku ho (Journal of East Asian Humanities)* 3, no. 1 (1941): 134-149.

Lee, James. "Migration and expansion in Chinese history." In *Human Migration: Patterns and Policies*, by William H. McNeill and Ruth S. Adams , 20-47. Bloomington: Indiana University Press, 1978.

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

Lee, James, and Cameron Campbell. "Social mobility from a kinship perspective: Rural Liaoning, 1789-1909." *International Review of Social History*, 2003: 1-26.

Lee, James, Cameron Campbell, and Guofu Tan. "Infanticide and family planning in rural Liaoning, 1774-1873." In *Chinese History in Economic Perspective*, by Lillian Li and Thomas Rawski, 149-176. Berkeley: University of California Press, 1992.

Lee, James, Cameron Campbell, and Lawrence Anthony. "A century of mortality in rural Liaoning, 1774-1873." In *Chinese Historical Microdemography*, by Stevan Harrell, 163-182. Berkeley: University of California Press, 1995.

Lee, James, Cameron Campbell, and Shuang Chen. *China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN) 1749-1909. User Guide.* Ann Arbor, MI: Inter-University Consortium for Political and Social Research, 2010.

Lee, James Z., and Cameron D. Campbell. China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN), 1749-1909. ICPSR27063-v7. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2012-11-21. doi:10.3886/ICPSR27063.v7.

Mare, Robert. "A Multigenerational View of Inequality." *Demography*, 2011.

Mare, Robert, and Xi Song. "Social Mobility in Multiple Generations." Hong Kong, China, 2012.

Oliver-Smith, Anthony, ed. *Involuntary Migration and Resettlement: The Problems and Responses of Dislocated People.* Boulder: Westview Press, 1982.

Ren, Yuxue. "

." 2012.

SCPTTJL. *Shuangchengpu tun tian ji lüe (Brief Records of the History of Shuangcheng State Farm).* Edited by Wang Lütai. Changchun: Jilin wen shi chu ban she, reprint 1990/1995 [1823].

Song, Xi, Cameron Campbell, and James Lee. "Descent Line Growth and Extinction from a Multigenerational Perspective." San Francisco, 2012.

Tsuya, Noriko, Feng Wang, George Alter, and James Lee. *Prudence and Pressure: Reproduction in Europe and Asia, 1700-1900.* Cambridge: MIT Press, 2010.

			7980	32
55	1866	1913	3284	40
			8113	32
56	1866	1913	3481	40
			9764	32
57	1866	1913	3720	40
			6757	32
58	1866	1913	3375	40
			8646	32
59	1866	1913	4212	40
			9415	32
60	1866	1913	2968	40
			7213	32
61	1866	1913	3458	40
			7101	32
62	1866	1913	3383	40
			8091	32
63	1866	1913	3061	40
			6858	32
64	1866	1912	11809	31
65	1866	1912	10051	31
66	1866	1912	7738	31
67	1866	1912	12640	31
			14	1
68	1866	1912	9670	31
			56	1
69	1866	1912	12855	31
			17	1
70	1866	1912	11368	31
71	1866	1912	8108	30
			24	1
72	1866	1912	8358	29
			12	1
73	1866	1912	11043	29
			9	1
74	1866	1912	12244	29
			38	1
75	1866	1912	10590	29
76	1866	1912	8059	29
77	1866	1912	9820	29
78	1866	1912	11110	29
79	1866	1912	12015	29
			23	1
80	1866	1912	7311	29
			33	1
81	1866	1912	7897	29

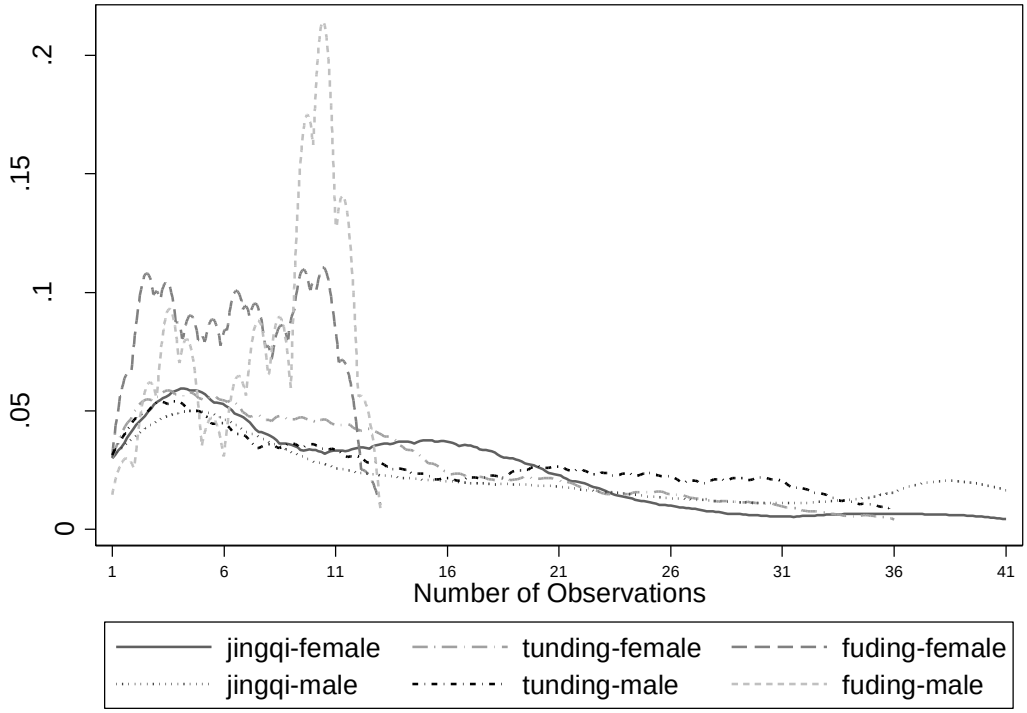


Figure A Distribution of Number of Observations per Individual
Source: CMGPD-SC, 1866-1913.
Note: Individuals with duplicated annual records are excluded.
[2] A trivial number of individuals who are missing on sex are also excluded.