Does Famine Have Long-Term Effects on Female Fecundity? Evidence from the Great Leap Forward Famine in China

Shige Song

January 20, 2010

ABSTRACT

Using data from a nationally representative sample survey in China, I identified a significant cohort difference in sterility between those who were born during the 1959-1961 Great Leap Forward Famine and those who were born immediately before and after. The increased risk of sterility for the famine cohort supports the claim that prenatal exposure to acute malnutrition may cause permanent damage to biological reproductive capacity, a research hypothesis that was first proposed over 10 years ago but have not been tested rigorously. The results of this research also contributes supportive evidence to the increasingly important idea of early developmental origins of health and disease and is potentially interesting to a broad scientific audiences as well as policy makers.

Women's reproductive capacity is influenced by many factors, both genetic and environmental (Lummaa, 2003). An important environmental factor is nutritional status (Menken et al., 1981; Bongaarts, 1980). While moderate chronic malnutrition has been shown to have little impact on female fecundity (John et al., 1987), actuate malnutrition, which is common in famine, may cause decrease in female fecundity because when food consumption falls below a critical minimum level women stop ovulating and therefor they cannot conceive (Bongaarts and Cain, 1981; Jowett, 1991). Fortunately, such a famine-induced decrease in female fecundity has been shown to be only temporary, and once the food consumption restores to normal levels, the fertility rates will recover quickly and there is no apparent permanent damage to the population's ability to reproduce (Stein, 1975; Stein and Susser, 1975).

Recent progress in identifying early developmental origins of adult health and diseases brings in an interesting new perspective. The "fetal origins hypothesis", first proposed by Barker (1992) and then tested and elaborated by many others (Gluckman and Hanson, 2006; Rasmussen, 2001), emphasizes the importance of prenatal nutritional condition in "programming" the fetus for the development of chronic disease in adulthood. It has been suggested that during fetal growth period, malnutrition can permanently influence the development of the fetus's organs undergoing critical growth by reducing the number of cells in the organs, which could be beneficial to the survival of the fetus later in life under the same poor nutritional conditions but will be detrimental under conditions of normal or over nutrition, and may lead to increased risk of health problems later in life (Lucas, 1991; Barker, 1995). Barker (1995) presents a case about how coronary heart disease is programmed in utero, which inspires Lumey and Stein (1997) to hypothesize that, severe malnutrition during famine may also affect the in utero development of fetus's organs responsible for production and regulation of female reproductive hormones, and therefore lead to permanently impaired female fecundity.

In this paper, I use self-reported pregnancy history information collected in a large national representative fertility survey in China and mixture logistic-lognormal survival model to test the

hypothesis that prenatal exposure to famine leads to reduced female fecundity by comparing reproductive performance of the cohort born during the Great Leap Forward Famine (1959-1961) in China against the cohorts born immediately before (1956-1958) and after (1962-1964). This paper represents the first comprehensive demographic research effort examining the relationship between prenatal exposure to severe malnutrition and biological reproductive capacity at adulthood, that fully integrates the century-long demographic literature on fecundity and sophisticated mixture statistical modeling methodology into a conceptual framework derived from the most recent development in biomedical science regarding early and developmental origins of health and diseases.

PRENATAL FAMINE EXPOSURE AND REPRODUCTIVE CAPACITY: THEORY AND EVIDENCE

The predicted cohort differences in reproductive capacity by Lumey and Stein (1997), that prenatal famine exposure leads to permanently impaired female fecundity, is not the only possibility. Also based on the concept of early developmental origins of health and diseases, Painter et al. (2008) proposed a competing hypothesis stating that although prenatal famine exposure may cause various kinds of chronic diseases at adulthood due to hormonal disregulations, as suggested by various researchers, the same disregulations may be at the same time associated with enhanced reproductive functions, because of developmental plasticity. Such a pattern is consistent with the theory of life history regulations, which proposes that fertility and body maintenance are mutually balanced, and that increases in investments in one are likely to be traded off by decreases in investments in the other (Stearns, 1992). A third possibility, selection effect by differential fertility and/or mortality, makes similar predictions to that by Painter et al. (2008): the famine cohort will have improved reproductive performance than the non-famine cohorts. But the reason is not related to early developmental origins but the general observation that famine survivors tend to be genetically well

endowed, which is likely to be associated with better reproductive performance, holding everything else equal.

Both Lumey and Stein (1997) and Painter et al. (2008) draw empirical evidence from the 1944-1945 Dutch Hunger Winer. Relying on hospital records and interview data of 1,116 women in the Dutch Famine Birth Cohort Study born during 1944-1946, Lumey and Stein (1997) report that famine exposure did not have significant effect on age at menarche, the proportion of having no children, age at first delivery, or the number of children; they did find an excess of perinatal deaths occurring among children of the famine-born women, providing some supportive evidence to their hypothesis that prenatal famine exposure reduces fecundity. Painter et al. (2008) also use hospital records combined with interview data from the Dutch Famine Birth Cohort Study, with a smaller sample size (473 women), and draw drastically different conclusion. According to Painter et al. (2008), women who experienced prenatal famine exposure have more children, with more twins, are less likely to remain childless, and start childbearing at a younger age. Painter et al. also tried to rule out the effect of differential fertility and mortality by comparing the observed characteristics of the mothers of the famine-born women against that of the mothers of the non-famine-born women. But since they cannot control for the unobservables, such as genetic endowments, their claims regarding selection effects remain inconclusive. A third study, also based on data from the Dutch Hunger Winter but not part of the Dutch Famine Birth Cohort Study, conducted by Elias et al. (2005), yields yet another different set of findings. Using information from 7,941 women born between 1932-1941, Elias et al. (2005) classified them into three mutually exclusive groups with respect to famine exposure – unexposed, moderately exposed, and severely exposed – and show that severe famine exposure during childhood significantly decreased chances of first and second childbirth at any given time after marriage or after the first childbirth, and increases the chances of having no or fewer children than desired due to medical reasons. This study, however, cannot be directly compared to the first two studies because its primary focus was on the effect of childhood famine exposure instead of prenatal famine exposure.

In trying to identify possible causes for the different results obtained from the same data sources, Lumey and Stein (2009) raise three possibilities: (1) the study sample used by Painter et al. (2008) are different from the study sample used by Lumey and Stein (1997) (sample size = 1,116 vs. 473); (2) they may have used slightly different famine exposure cutoff standard; and (3) results based on women aged 43 (when they were interviewed by Lumey and Stein) maybe genuinely different from results based on women aged 58 (when they were interviewed by Painter et al.). In addition to these apparent reasons, there may be more fundamental issues underneath the Lumey-Painter disagreement, something that is rooted in their conceptual ambiguities and analytical weaknesses.

Some of the findings by Lumey and Stain and Painter et al. cannot be directly compared because they used different outcome variables; among those that can be directly compared, the differences between Lumey and Stein and Painter et al. focus on the following outcomes: (1) childlessness; (2) birth intervals; and (3) number of children. These outcomes have been the central focus of demography for decades and has generated a long line of research literature (Gini, 1924; Henry, 1964; Trussell and Wilson, 1985; Larsen and Menken, 1989; Wood et al., 1994; Leridon, 2008), with special attention devoted to measurement and analytical issues (Bongaarts, 1975; Heckman and Walker, 1990; Larsen and Menken, 1989; Weinstein and Stark, 1994; Wood et al., 1994; Dunson and Zhou, 2000). I begin my analysis with a comprehensive review of this literature, aiming to uncover the facts underlying the differences between the results by Lumey and Stein (1997) and Painter et al. (2008).

MEASURING FECUNDITY: A DEMOGRAPHIC APPROACH

Fecundity is closely related to fertility: while fecundity refers to the biological capacity to reproduce, fertility refers to the realization of such capacity, the actual production of the new generations. Both fertility and fecundity are important demographic concepts and have been studied

intensively. Since fecundity is a latent theoretical construct that cannot be directly measured, demographers have to make inference about fecundity based on fertility; and for this reason, they rely on sterility and fecundability in empirical research. In this section, I review the demographic literature on the measurement and analytical issues with respect to fecundity, with special focuses on the following aspects: (1) The validity of treating obtained level of fertility as indicator of fecundity; (2) The validity of estimating and using fecundability based on time to pregnancy/birth information in regulated fertility populations, that is, population in which contractive measures are widely used; (3) The validity of estimating and using sterility based on time to pregnancy/birth information in regulated fertility population with unknown fertility preference; and (4) How to achieve greater validity in a new study.

Obtained Fertility Is Not A Good Indicator of Fecundity

Fertility of a population is determined by fecundity and a number of fertility-inhibiting factors such as late marriage and marriage disruption, deliberate birth control, and breast-feeding (Bongaarts, 1980). Fertility can be treated as an indicator of fecundity only in populations with "natural fertility", where contraceptive use is completely absent or negligible. The Dutch Hunger birth cohorts (individuals born around 1944-1945) can hardly be characterized as a natural fertility population: when they reach childbirth age, in the 1960s, the vast majority of them used highly effective contraceptive measures (Leridon, 1981; Moors, 1978). In this situation, the total number of children each woman ended up having depends much more on personal choice and social norms than biology, and should not be used as indicator for biological reproductive capacity.

In that sense, the difference in findings between Lumey and Stein (1997) and Painter et al. (2008) with respect to the cohort difference in the total number of children can be caused by many factors and is largely irrelevant to their central research objective: testing to see prenatal famine exposure has impacts on biological reproductive capacity. Due to the inherently complicated nature

of the issue and the difficulties in gaining adequate control, the use of total number of children as indicator of fecundity should be discouraged in future research on contemporary populations in which contraception has been widely practiced.

Estimating and Comparing Fecundability in Populations with Regulated Fertility

Reproductive capacity, or fecundity, is inherently unobservable and cannot be directly measured. Following Gini (1924), demographers instead use fecundability, the monthly probability that a woman in a susceptible state will conceive if she is exposed to unprotected sex intercourse, as their primary measure of fecundity. Fecundability is usually estimated using data on duration between time of marriage and time of first conception (total fecundability) or first birth (effective fecundability) to avoid the impact of postpartum amenorrhea and breast-feeding on measured times to conceptions that affects data on higher-order conceptions. There are cases where it is advantageous to use multiple birth intervals in estimating fecundability because information on first birth interval is biased by the presence of premarital pregnancy/birth or marriage is itself selective with respect to fecundability (Wood, 1994, Pp.286).

Estimating fecundability is straightforward in populations with natural fertility where contraceptive use is negligible; it is not the case, however, with most contemporary populations where contraception is widely used. This is why existing studies on fecundability and sterility are mostly based on historical populations or special populations where contraceptive use is strictly prohibited (Larsen and Vaupel, 1993; Wood et al., 1994). Only when information on exactly when the contraceptives were used are available, it is possible to get unbiased estimate of fecundability and sterility of contemporary populations with controlled fertility, by excluding those intervals with contraceptive use from the periods of genuine exposure to the risk of conception (Wood, 1994).

Since the Dutch Hunger birth cohorts do not qualify as natural fertility population, and neither Lumey and Stein (1997) nor Painter et al. (2008) incorporated information on contractive use in

their analysis, their results on cohort difference in the timing of childbirth should be interpreted with great caution and cannot be taken as sound evidence for the effect of prenatal famine exposure on reproductive capacity. Furthermore, since fecundability is influenced by both biological and behavioral factors (Weinstein and Stark, 1994), without adequately controlling for behavioral difference between cohorts, it is much less convincing to make claims about the difference in biological reproductive capacity between the famine and the non-famine cohort based on the estimated cohort difference in fecundability.

Estimating and Comparing Sterility in Populations when Fertility Preference is Unknown

Since fecundability, by definition, only applies to fecund women, a second term, sterility, is then used to delineate whether a woman is biologically capable of conceiving. Unlike fecundability, which is influenced by both biological and behavioral factors, sterility is influenced by biological factors only (Wood, 1994; Weinstein and Stark, 1994). In this sense, sterility is a better indicator than fecundability of human reproductive capacity if the primary focus is the biological aspect of it, as is the case for both Lumey and Stein (1997) and Painter et al. (2008), and the present research as well. [1]

"Childlessness" does not, however, necessarily mean sterility. For example, a childless woman can be perfectly capable of childbearing, and the only reason she ends up being childless is because she does not want any children, and the increasing availability of highly effective modern contraceptive technology makes it possible to achieve this goal. In other words, to make inference about sterility based on the observed childlessness, it is important to distinguish between voluntary and involuntary childlessness. When the Dutch Hunger birth cohort reached reproductive age, up to one fourth of them chose not to have any children (Den Bandt, 1980). Without knowing the

[1] I use the term sterility to refer to what is otherwise called "permanent" "complete", and "primary" sterility: an irreversible biological process that begins before any offsprings have produced and continue until the end of the reproductive period (Wood, 1994, Pp.443-444).

fertility preference of each woman in the selected sample, it is not possible to separate this group of voluntary childless women from those who lack the biological reproductive capacity. Given the sheer size of this special group, its cohort distribution can easily dominate the overall cohort pattern of the estimated childlessness and thus lead to potentially erroneous conclusion about the effect of prenatal famine exposure on female sterility at adulthood.

In short, although sterility is a good indicator of fecundity, since neither Lumey and Stein (1997) nor Painter et al. (2008) adequately controlled for fertility preference and excluded voluntarily childless women from their cohort comparison, their results on cohort difference in childlessness should not be interpreted as direct evidence for the impact of prenatal famine exposure on biological reproductive capacity.

Other Issues and Suggestions

Besides the three major issues discussed above, there are several other issues in the demographic literature that are relevant to the research effort to assess the potential effect of prenatal famine exposure on fecundity at adulthood.

The estimation of fecundability and sterility has been treated in demographic literature as separate issues with each has its own dedicated literature (Larsen and Menken, 1989; Trussell and Wilson, 1985; Weinstein et al., 1990; Larsen and Vaupel, 1993; Kallan and Udry, 1986). During the past 20 years or so, demographers began to realize that sterility and fecundability, as two different aspects of fecundity, should be modeled jointly to reduce bias and increase efficiency (Wood et al., 1994). Various statistical estimation procedures have been proposed to simultaneously estimate sterility and fecundability, with different underlying behavioral assumptions and varying degree of methodological sophistication and technical complexity (Heckman and Walker, 1990; Dunson and Zhou, 2000; Wood et al., 1994; Larsen and Vaupel, 1993), so applied researchers can choose the one that suites their research questions best.

Both time to pregnancy and time to first birth have been used in the demographic literature to estimate fecundability. The former method yields what is called "total fecundability" while the latter yields what is called "effective fecundability". Due to the difficulties in detecting early pregnancies, especially when early pregnancy ends in fetal loss during the first few days of gestation, most empirical studies end up estimating "apparent fecundability", fecundability estimated based on time to the first identified pregnancy, instead of total fecundability. Since childbirth is much more tightly regulated by non-biological factors such as social norms, religion, financial constraints, and state policy (in the case of China), apparent fecundability may be a more sensitive indicator than effective fecundability of the prenatal famine exposure effect. If possible, both should be carefully investigated.

In summary, a careful review of the demographic literature on human fecundity leads to the following suggestions in conducting cohort comparison to assess the potential impact of prenatal famine exposure on fecundity. First of all, total number of children is not a good indicator for fecundity. Second, it is important to control for contraceptive use and fertility preference when estimating and comparing fecundability and sterility based on time to conception/birth data in contemporary populations with regulated fertility. Third, sterility may be a better indicator of the biological reproductive capacity than fecundability because it is much less influenced by behavioral factors. Fourth, both fecundability and sterility should be estimated simultaneously to reduce bias and improve efficiency.

THE CURRENT STUDY

The current study aims to assess the potential effect of prenatal exposure to the 1959-1961 Great Leap Forward Famine in China on fecundity by comparing the level of sterility between three Chinese birth cohorts: those born in 1956-1958 (the pre-famine cohort), those born in 1959-1961 (the famine cohort), and those born in 1962-1964 (the post-famine cohort). This study differs from

previous studies in several important ways. First of all, I took advantage of the sophisticated demographic conceptual framework toward human fecundity to keep the analysis focused. Second, by focusing on a carefully selected subsample of rural residents who wanted to have children and did not use any contraception before their first pregnancy/birth, I was able to minimize the potential bias caused by the unobserved heterogeneity in fertility preference and behaviors. Third, I focused on the Great Leap Forward Famine in China, which is the most severe famines in modern human history with respect to duration, affected population size, and the severity of malnutrition (as reflected by the number of excess mortality), making it an ideal test case for long-term health and demographic consequences of prenatal exposure to malnutrition. Last but not the least, utilizing large national representative sample survey data guarantees sufficient statistical power to detect changes and differences in relatively rare events such as the permanent loss of biological capacity to reproduce even before childbirth age.

The Great Leap Forward Famine

There have been a number of studies on the causes and the magnitude of the 1959-1961 Great Leap Forward Famine in China (Peng, 1987; Ashton et al., 1984; Yao, 1999; Kung and Lin, 2003). These studies revealed several important facts about the famine. First of all, the Great Leap Forward Famine, which started in the early 1959 and ended in 1961 in most part of the country, caused about 30 million excess mortality and 33 million fetal loss, making it one of the most disastrous humanitarian tragedies in human history. Second, the famine was much more severe in the rural areas than in the urban areas. Reliable data on changes in daily calorie intake during that period of time is unavailable, indirect estimates suggest that although grain availability declined in both urban and rural areas during the famine years, the decline was much smaller in urban areas than in rural areas, which exacerbated the preexisting urban-rural difference in available food per capita, and push the daily calorie intake of many rural residents below the subsistence level during the

famine years (Peng, 1987; Lin and Yang, 2000; Song et al., 2009). By contrast, the urban population was much better protected, and starvation was rarely heard among the urban population, and there is no sign of suddenly increased mortality, even among the most fragile ones: newly born infants (Song et al., 2009). Third, the population recovered from the famine quickly: in 1962, right after the famine was over in most provinces in China, mortality returned to the pre-famine level and fertility increased to a level that was even higher than the pre-famine level, and continued for several years (Ashton et al., 1984; Peng, 1987), caused by typical compensatory fertility behavior after famine (Bongaarts and Cain, 1981).

An important reason for this urban-rural difference in the famine severity is the differential treatment urban and rural population received in the Chinese centrally planned economic system: while urban residents are granted legal right to food security through the system of food rationing, rural residents, as the food producers, were burdened with legal obligations to hand in their product with coercive quotas through the national food procurement system and were entitled only to the residual food supply (Lin and Yang, 2000). In the years of bad harvest (such as 1959-1961), there would not be much residual food supply left for the peasants themselves after finishing the food procurement quota, and the extreme measures used by the state government to extract "surplus" food from the peasants gave the peasants no choice but to comply, although resistance and conflicts were not unheard of (Walker, 1984).

In other words, underneath the seeming urban-rural difference in famine excess mortality lies a particular type of institutional setting combining a food procurement system (to extract "surplus" food from peasants), a food rationing system (to guarantee food safety for urban residents), and a household registration system (to control the rural-to-urban migration), designed to achieve rapid industrialization at the cost of agriculture and the interest of rural population. In a time of food shortage, such a system would sacrifice the food entitlement of the rural population to protect the food entitlement of the urban population, and this is exactly what happened during 1959-1961 in China.

I focus primarily on the rural population in this study.

The One-Child Policy and Changing Fertility Behaviors in China

When the selected cohorts (1956-1964) entered childbearing age, China was during the transition from a "later, longer, fewer" family planning policy to a much more rigid one-child policy. While the later, longer, fewer policy encourages people to have no more than two children, late marriage, and long birth interval, the one-child policy prohibited any higher-order birth while relaxed the late marriage and long birth interval regulations. These policy changes have important implications to the design and analysis of the present study.

First of all, the cross-cohort decline in marriage age and birth interval are largely policy-driven and should not be taken as evidence of the effect of prenatal famine exposure on the biological reproductive capacity, a strong substantive reason not to rely on fecundability as the primary outcome of interest in the context of Chins. Second, neither the "later, longer, fewer" policy nor the one-child policy can be characterized as anti-natal because they all consider having a certain number of children for each married couple as natural and necessary and should be encouraged. What they aim to prevent is to have "too many" children. Although the exact definition of "too many" changes over time, but having one child has never been considered "too many" under any family planning policies in China. This is a strong substantive reason to focus on the first pregnancy/birth instead of all pregnancy/birth in modeling time-to-pregnancy/birth data, which can avoid potential confounding effect of the changing family planning policy.

A third issue is contraceptive use. Although contractive use is generally prevalent among Chinese population, partly because of the family planning policy, very few women practiced contraception before the first birth, especially among rural population (Choe and Tsuya, 1991; Short et al., 2000).

RESEARCH DESIGN

Data and Sample

I use data from the National Family Planning and Reproductive Health Survey (NFPRHS), a nationally representative sample survey conducted by the State Family Planning Commission of China (SFPC) in 2001. The survey utilized a stratified multistage clustered sample to collect information from 39,586 women (29,512 rural and 10,074 urban) aged 15 to 49 and living in family households in 31 provincial administrative units in China. All selected women were asked to provide a complete pregnancy history, including the year and month of pregnancy termination and the outcome of each pregnancy.

The following key variables are used in the present analysis.

Ever pregnant. This is a binary variable, constructed from the pregnancy history roster, indicating whether a respondent has ever been pregnant in her life.

Time to first pregnancy. This is a continuous variable measuring the duration between the first marriage and the first pregnancy, for those who have ever had any pregnancies.

Live birth. This is a binary variable, constructed from the pregnancy history roster, indicting whether a respondent has ever had any childbirths in her life.

Time to first live birth. This is a continuous variable measuring the duration between the first marriage and the first childbirth, for those who have ever had any live births.

Birth cohort. The famine cohort includes those who were born in 1959-1961, the pre-famine cohort includes those who were born in 1956-1958, and the post-famine cohort includes those who were born in 1962-1964. Alternative cohort definitions have been experimented as part of the sensitivity analysis. For example, the pre-famine cohort is defined as those born in 1955-1957 and the post-famine cohort is defined as those born in 1963-1965, leaving those born in 1958 and those born in 1962 out of the analysis. The results turn out to be robust to different cohort definitions,

and the main substantive findings remain unchanged.

Educational attainment. This is a four-category ordinal variable measuring respondent's educational attainment. The four categories are: no schooling at all, primary school education, junior high school education, and senior high school education and above.

Ethnicity. This is a generated variable based on the original ethnicity variable in the survey by classifying all non-Han ethnic minorities into one category.

Fertility preference. This three-category variable is generated from the original variable "what do you think the ideal number of children in a family is?" I classify any numerical answer other than "0" as "want children", the answer "0" and "does not matter" as "do not necessarily want children". I include only those who want to have children in the analysis.

Contraceptive use. By utilizing information on the timing of first pregnancy/birth and the timing of first contraceptive use, this three-category variable captures the sequential order of first pregnancy/birth and first contraceptive use. The three categories are: (1) did not use contraceptive measures before first pregnancy/birth, (2) used contraceptive measures before the first pregnancy/birth, and (3) ordering cannot be determined. I excluded both (2) and (3) from the analysis to make sure that the analytical sample can be treated as sample with natural fertility.

Table 1 reports urban-rural difference in fertility preference and the use of contraceptive measures. No rural women claimed that they don't want any children; by contrast, 3% urban women made the same claim. On the other hand, 6% of urban women used contractive measures before their first birth, by contrast, only 1% of rural women did so. This urban-rural contrast provides additional justification to exclude urban women from the analytical sample.

Table 2 reports descriptive statistics of the analytical sample, by birth cohort. The sample includes rural women who want to have children and who have not used any contraceptive measure before their first pregnancy (in the time-to-pregnancy analysis) or birth (in the time-to-birth analysis).

Statistical Methodology

At the time of the survey interview in 2001, individuals in the selected cohorts were aged 37-45, the age range during which most Chinese women have completed childbirth, but have not lost biological reproductive capacity (menopause) (Lavely, 1986). The survey data have information on whether and when each sampled woman got pregnant or have a child up until 2001 but not beyond that time. The older cohorts will almost certainly show higher pregnancy/birth rate than younger cohorts simply because they have had longer time to fully realize their biological reproductive potential. Since ordinary logistic regression cannot handle right censoring, it produces biased estimates of cohort difference.

Survival analysis model provides a good alternative to handle right-censored data without information loss. Most survival models, however, assume that censoring is independent of the event, and that everyone will eventually experience the event of interest. If a subset of the population will never experience the event because they, for some reasons, have never been "at risk", the assumption no longer holds and will lead to biased estimates because all the individuals that are not at risk fall into the censored group. In the context of fecundity study, people who are sterile are not at risk of getting pregnant or having childbirth, excluding them from the analytical sample or including them in the analytical sample without special treatment leads to biased estimates of fecundability, a point that has been discussed extensively in the past (Heckman and Walker, 1990; Wood et al., 1994; Dunson and Zhou, 2000). In the present case, since sterility plays such a vital role in research design, the fact that survival model cannot directly estimate the prevalence and variation of sterility is even more problematic.

I tackled this issue by jointly estimating a logistic regression on sterility status and a parametric log-normal survival model on the timing of first pregnancy/birth, known as the "cure model", "long-term survivor model", or "split-population" in the survival analysis literature (Farewell, 1982; Maller and Zhou, 1996).

Following Sposto (2002) and Lambert et al. (2009), the survival function of a mixture survival model can be defined as:

$$S(t) = \pi + (1-\pi)(1-F(t)) \tag{1}$$

where π represents the fraction women who are sterile, and $F(t)$ denotes a statistical distribution function. The hazard function can be written as:

$$h(t) = \frac{(1-\pi)f(t)}{S(t)} \tag{2}$$

where $f(t)$ is the density function of $F(t)$. Following Wood et al. (1994), I choose log-normal distribution for $F(t)$, which makes sense in the case of time-to-pregnancy/birth, because it is known to have an inverse J-shaped hazard function.

$$F(t) = \Phi(log[\lambda t]^{\gamma}) \tag{3}$$

where $\Phi(\cdot)$ is the standard normal distribution function.

Covariates can be introduced into both the logistic regression for sterility π and the log-logistic survival function $F(t)$. For the sterility equation, logistic transformation has been the most common choice and the coefficients have an odds ratio interpretation:

$$log\frac{\pi(x)}{1-\pi(x)} = \alpha + \beta X \tag{4}$$

For the log-logistic survival model, covariates can be introduced to influence either the scale parameter λ or the shape parameter γ or both. In the main analysis, I restricted the shape parameter

to be constant and only allow the scale parameter to vary with covariates [2]:

$$log(\lambda) = \alpha + \beta X + \mu \tag{5}$$

where μ follows the standard normal distribution with mean 0 and fixed standard deviation. Since log-normal model is not a member of proportional hazard model, its coefficients have a time ratio interpretation: one unit increase in covariate X produces $exp(\beta)$ unit of change in the time-to-event.

More technical details of mixture survival model, including model likelihood function and numerical maximization details can be found in Sposto (2002) and Lambert et al. (2009). The analysis relies on user-contributed mixture analysis modules for Stata that implements the above descriptions (Lambert, 2007).

ANALYSIS

The analysis is conducted in four steps. I first present descriptive analysis on cohort trend in the proportion of women in the analytical sample who have never been pregnant or who have never had live births. In the second step, I present results from mixture logistic-lognormal models on time to first pregnancy. Then I present results from mixture logistic-lognormal models on time to first birth. At the last step, I present results from mixture logistic-lognormal models on both time to first pregnancy and time to first birth for urban population, as comparison to the main analysis based on rural population.

[2] In sensitivity analysis, I also allowed the shape parameter to vary with covariates. Introducing covariates into the shape parameter, however, does not change parameter estimations in the sterility model or in the fecundability model.

Descriptive Analysis

Figure 1 shows the cohort trend in the proportion of rural women who have not used any contraceptive measures before the first pregnancy/birth, who want to have at least one child, and yet have never been pregnant and the proportion of women who have never had live childbirth. Despite some fluctuations, the effect of prenatal famine exposure on both conception and childbirth is quite clear: the proportions of childless and never pregnant rural women rise sharply from the 1958 to the 1959 cohort, reach their peak value in the 1960 birth cohort, then start to fall until hit the floor in the 1962 birth cohort. Since the Great Leap Forward Famine began in most areas in early 1959 and ended in most areas in 1961, the empirical pattern agrees with the claim that prenatal famine exposure reduces fecundity.

Interestingly, the line representing childbirth and the line representing conception overlap in some cohorts, indicating that the first pregnancy and the first childbirth happen at the same time for these women. The biggest discrepancy occurred at the 1960 cohort, of which about 2% never had a child, while only 1.5% never got pregnant. This suggests that one aspect of the negative impact of prenatal famine exposure on female fecundity is to increase the risk of involuntary fetal losses, a topic that deserves further investigation.

Time to First Conception

Table 3 reports results from three mixture logistic-lognormal model on time to first pregnancy. For each model, the output is divided into two panels: a logistic regression predicting sterility, controlling for right censoring; and a log-normal accelerated failure time model estimating fecundability, for those who were at risk.

Model 1 is the baseline model that includes birth cohorts as the only covariates in both the sterility and the fecundability equation. Model 2 adds two more covariates, educational attainment

and ethnicity, into both the sterility and the fecundability equation of Model 1, to see if adding these covariates makes any changes to the estimated cohort difference. Model 2 fits the data less well than Model 2, based on both *AIC* (54807.3 vs. 54845.0) and *BIC* (54910.9 vs. 54893.4), indicating some of the newly included covariates are redundant. Model 3 includes only those significant covariates in Model 2, and thus represents the best fitted model, as indicated by both *AIC* and *BIC*.

The estimated cohort difference in both sterility and fecundability are robust to alternative model specifications. In all three models, (1) both the pre-famine and post-famine cohorts have much lower risk of sterility than the famine cohort, and (2) there seems to be a secular trend for shorter interval between marriage and first pregnancy/birth, which is consistent with past research (Wang and Yang, 1996; Hong, 2006), even though the difference between the pre-famine and the famine cohorts are not statistically significant.

Based on the best fitted model, the risk of sterility does not depend on socioeconomic factors such as education and ethnicity; and the only important determinant is birth cohort, which captures differential prenatal famine exposure. This is an important finding, for several reasons. First of all, since fertility preference may be influenced by many factors and may change over time, it is difficult to obtain adequate measure of fertility preference in survey research. Even with the rigid sample selection criteria that excluded urban women, women who either did not want children or did not care about having children or not, and women who used contraception before first pregnancy/birth, there is still not guarantee that every woman included in the analytical sample wanted to have children and the only reason some of them ended up being childless was that they lack the biological capacity to reproduce. The fact that the estimated sterility does not have a SES gradient in the selected sample adds a great deal of confidence to the validity of this crucial assumptions, on which the main substantive conclusions depends. [3]

[3] As part of the sensitivity test, I also included a set of dummy variables for provinces in the model; the results show that province does not have significant influence on sterility, and adding these dummy variables has little impact on the estimated cohort difference.

Fecundability, on the other hand, is shown to be influenced by birth cohort, educational attainment, and ethnicity. As discussed before, unlikely sterility, which is influenced by biological factors only, fecundability has both a biological and a behavioral component (frequency of sexual intercourse). Unless one adequately can control for the intercourse frequency (which is difficult, if not possible in social surveys), the estimated cohort fecundability cannot be interpreted as representing the effect of prenatal famine exposure on biological capacity of reproduction. The presence of significant SES gradient in fecundability supports this argument: since fertility behavior is influenced by socioeconomic factors (Bongaarts, 1978), without adequately controlling for the behavioral heterogeneity, some of its effect will be forced upon socioeconomic factors that have influences on such behaviors. One would naturally conclude that once the behavioral heterogeneity can be adequately controlled for, the effect of SES on fecundability will disappear. This is an interesting research hypothesis for future study to test, when better data is made available.

In summary, the main results reported in Table 3, that the famine-born cohort on average face 50% higher odds than the pre-famine cohort and 62% higher odds than the post-famine cohort of sterility, estimated from time to first pregnancy, provides supportive evidence to the claim that prenatal famine exposure has negative impact on female fecundity.

Time to First Birth

Analysis based on time to first birth analysis information, as reported in Table 4, yields results that are highly consistent with analysis based on time to first pregnancy information, as reported in Table 3. Model comparison based on both *AIC* and *BIC* indicates Model 3, in which the sterility is modeled as function of birth cohort only while the fecundability is modeled as function of birth cohort and socioeconomic indicators such as education and ethnicity yielded the best fit. Famine cohort suffers from significantly higher prevalence of sterility, compared to both the pre-famine and the post-famine cohorts, supporting the claim that prenatal famine exposure has negative im-

pact on female fecundity. Cohort pattern in changes in fecundability suggests a secular trend in shortening interval between marriage and first birth, which is consistent to earlier studies. The finding that better educated and Han ethnic majority women tend to have shorter birth interval (higher fecundability) is also consistent with previous research (Wang and Yang, 1996), and may be result of the fact that better educated women are more likely to meet and find their own spouse-to-be, which are associated with greater intimacy and increased sexual activity. Similar findings have been discovered in other Asian societies undergoing the process of industrialization (Rindfuss and Morgan, 1983).

In summary, results reported in Table 3 and 4 provides clear evidence supporting the argument that prenatal exposure to severe famine may lead to impaired biological reproductive capacity, which was suggested by Lumey and Stein (1997) but was not well supported by their own empirical study based on data from the Dutch Hunger Famine.

Figure 2 shows predicted prevalence of sterility for the three cohorts included in the analysis, based on both time-to-pregnancy and time-to-birth information. Several patterns worth discussion. First of all, the prevalence of sterility is low for the pre-famine cohort (about 0.7%), and there is little difference in the prevalence of sterility based on time-to-pregnancy and that based on time-to-birth, indicating that almost everybody's first pregnancy leads to live birth in the pre-famine cohort. Second, the prevalence of sterility rises sharply from the pre-famine to the famine cohort, from about 0.7% to 1.4% (based on time-to-pregnancy) or 1.5% (based on time-to-birth), an increase of over 100%. Also, the pattern based on time-to-pregnancy and that based on time-to-birth becomes more noticeable, indicating that an increasing number of women's first birth did not lead to live birth in this cohort. Third, the prevalence of sterility declines drastically from the famine cohort to the post-famine cohort, from 1.4-1.5% to 0.5-0.6%, a decline of 60%. The gap between time-to-pregnancy and time-to-birth narrowed somewhat among the post-famine cohort, but not as low as the pre-famine level.

Urban Trends

As comparison to the main findings reported in Table 3 and 4, Table 5 reports results from two mixture logistic-lognormal models on time to pregnancy and two models on time to first birth for urban population. As discussed earlier, the key differences between the urban and rural population includes: (1) the famine was much less severe in urban areas than in the rural areas, partly because the state socialist redistribution system guarantee some level of food security to the urban residents but not to the rural residents; (2) the urban population are much more heterogeneous with respect to fertility preference and fertility behaviors, which may contaminate the main substantive findings.

Comparing Pregnancy Model 1 and 2 and Birth Model 1 and 2 shows that including birth cohort in either the sterility or fecundability model reduces model fit, suggesting no birth cohort variation among the urban population with respect to both sterility and fecundability. In fact, as these models show, sterility among urban population does not depend on any of the covariates considered in the research; and fecundability only depends on education and ethnicity. In other words, sterility is a random biological phenomenon among the urban population, as in most populations under normal conditions (Wood, 1994, Pp.445-448).

Due to the small urban sample size and the rarity of sterility (nine cases for the pre-famine cohort, 5 cases for the famine cohort, and 14 cases for the post-famine cohort), the possibility that the lack of statistical significance in cohort difference in sterility may be partly caused by insufficient number of sterility cases cannot be ruled out. A more definitive answer to the question why there is no cohort variation in sterility among the urban population can only be obtained when better data with larger sample is made available in the future.

DISCUSSION AND CONCLUSION

Using a carefully selected subsample of rural women who wanted to have children and who did not use contraceptive measures before their first pregnancy/birth from a large national representative sample survey conducted in 2001 in China, this study reveals significant cohort difference in the risk of sterility – the lack of biological capacity to reproduce. To be more specific, the cohort born during the Great Forward Famine in 1959-1961 are facing over 60% higher odds than the cohort born immediately after and about 50% higher odds than the cohort born immediately before of being sterile even before entering childbearing age. Analysis using information on time to first pregnancy and that using information on time to first birth yield highly consistent results, and strongly support the claim that prenatal famine exposure has negative effect on female fecundity. The meaning of cohort difference in fecundability, the probability that a fecund woman will conceive after a month of unprotected sexual intercourse, is much more complicated. In the Chinese context, the difference in fecundability between the cohort born in 1956-1958, 1959-1961, and 1962-1964 are mostly driven by the shift in state family planning policy, from the "later, fewer, longer" policy to the one-child policy, and the secular trend in increasing educational attainment and non-family employment (Wang and Yang, 1996).

To isolate meaningful cohort comparison, I made several important design choices. First of all, I focus primarily on rural population because the famine was much more severe among rural population than among the urban population, which was produced and exacerbated by the centrally planned economic system along with other institutional arrangements at the time; and because the rural population behave more likely the population with natural fertility, at least with their first contraception/birth: almost all women want to have children, and virtually nobody used contraceptive measures before their first pregnancy/birth. Second, I further excluded those who either stated that having at least one child is not very important or those who either used contraceptive before their first pregnancy/birth or cannot be sure whether they used contraceptive before their

first pregnancy/birth. The three exclusion criteria together can guarantee the cohort difference in time to pregnancy/birth is largely free of confounding effect of differential fertility preference and behavior. Third, the conceptual separation between sterility and fecundability adds further control of the potential confounding effect of differential fertility behaviors. Following a long line of demographic literature, I made the argument that cohort difference in sterility reflects mostly cohort difference in the biological reproductive capacity, while cohort difference in fecundability reflects a mixture of cohort difference in biological reproductive capacity, changes in behavior that are driven by state policy, and changes in behaviors that are driven by secular trend in the increase in education and employment. In short, with adequate control for fertility preference and behaviors, cohort difference in sterility can be a good indicator of the potential effect of prenatal famine exposure on fecundity; while cohort difference in fecundability is not. Fourth, I choose to focus on the first pregnancy/birth only to avoid the potentially serious confounding effect of family planning policy, including the "later, fewer, longer" policy in the 1970s and the one-child policy since the early 1980s.

The urban-rural difference in cohort pattern in sterility worth more discussion. Clearly, both urban and rural populations suffered from the three-year famine to some extent. The most important reason why there is no sign of negative effect of prenatal famine exposure on female fecundity among urban women may be the simple fact that the daily per capita grain availability is much higher among urban population than among rural population; the difference is so big that in many provinces the daily available grain per capita for the urban population during the famine time is higher than the daily available grain per capita for the rural population during the non-famine time (Peng, 1987, Table 4). In other words, it is very likely that the reason why the urban famine cohort did not suffer from impaired fecundity was because (1) only prenatal exposure to acute malnutrition impairs fecundity, and (2) the urban population did not suffer from such acute prenatal malnutrition because of the protected food safety provided by the centrally planned economy that is not available to rural population. This is the topic on which more extensive interdisciplinary

research effort is required to get a more definitive answer.

One would naturally wonder, in the context of famine, whether and to what extent selection by differential mortality and fertility may have influenced the results. After all, the extant famine cohort constitutes only a fraction of the original cohort, and selection effect has been proven to be very important determinant in shaping up the observed cohort in child and adolescent mortality and schizophrenia (Song et al., 2009; Song, 2008, 2009). The fact that my result shows famine cohort has much higher sterility prevalence than the non-famine cohorts can only means one of two things: (1) it is possible that the estimated cohort pattern (famine cohort about 50% higher odds of sterility than non-famine cohort) represents the lower bound of the true effect of prenatal famine exposure on fecundity because the famine survivors are likely to be genetically well endowed and should have higher-than-average reproductive capacity; or (2) the estimated cohort pattern represents the true effect of prenatal famine exposure and factors that influence reproductive capacity is unrelated to the factors that determine fertility and mortality during the famine. In either case, the results presented in here provides strong evidence that support the claim that prenatal famine exposure permanently impairs female fecundity.

Since sterility is a relatively rare phenomenon, some may have reservations about the practical importance and policy relevance of the present research. In many societies, sterility (and the resulted involuntary childlessness) is stigmatized and caused a great deal of pain and harm to the women (Miall, 1985; Whiteford and Gonzalez, 1995). Since sterility does not have a clearly defined etiology (other than sexually transmitted diseases), it is often considered as the women's own fault (sins and wrongdoings, etc.). The social pressure is even greater in societies with strong Confucian pro-natal cultures such as China. This research demonstrated, for the first time, that over .17 million sterile women out of the 16 million rural women born in 1959-1961 were caused by prenatal exposure to acute malnutrition. [4] For many people, the Great Leap Forward Famine was over long time ago, before even they can remember anything; but for theses .17 million rural

[4]Total number of female birth in 1959-1961 were calculated from Table 2 in Lin and Yang (2000).

women who have suffered and struggled, the tragedy was never over and they have been living their whole life in its giant shadow without even knowing. As they are approaching retirement age soon, what is waiting for them is the insurmountable difficulty of old age support because the predominant form of old age support in rural China is still family-based: without children, they have nothing; and their only hope is the state policy that does not exist yet. It is time for changes.

The idea that health and disease may have an early life origins is one of the most important innovations in medicine and epidemiology in recent years (Rasmussen, 2001; Gluckman and Hanson, 2006). Since health plays such a pivotal role in virtually every aspect of social life, it has made its way into the mainstreams of many social science disciplines. By focusing on the potential impact of prenatal famine exposure on fecundity, a key concept of demography, I was able to improve upon existing studies to achieve more accurate and meaningful cohort comparisons by controlling for population heterogeneity in fertility preference and behaviors, something that have not been done in past research, which explains the confusing and even contradicting findings. One potential weakness of this research is the reliance on urban-rural distinction as a proxy measure the severity of malnutrition, which works reasonably well but cannot be used to infer important information such as: how severe the malnutrition has to be to cause permanent damage to female reproductive function? A better strategy is to use retrospective information on dietary intake during the famine years for the selected individuals through personal interview. Since such information are not available any existing data sources, new data collection efforts will be required.

REFERENCES

Ashton, B., K. Hill, A. Piazza, and R. Zeitz. 1984. "Famine in China, 1958-61." *Population and Development Review* 10:613–645.

Barker, D. J. P. 1992. "The Fetal Origins of Diseases of Old Age." *European Journal of Clinical Nutrition* 46:S3–9.

Barker, D. J. P. 1995. "Fetal origins of coronary heart disease." *British Medical Journal* 311:171.

Bongaarts, J. 1975. "A method for the estimation of fecundability." *Demography* 12:645–660.

Bongaarts, J. 1978. "A framework for analyzing the proximate determinants of fertility." *Population and development review* 4:105–132.

Bongaarts, J. 1980. "Does malnutrition affect fecundity? A summary of evidence." *Science* 208:564–569.

Bongaarts, J. and M. Cain. 1981. "Demographic response to famine." In *Famine*, edited by K. Cahill, pp. 44–59. New York: Orbis.

Choe, M.K. and N.O. Tsuya. 1991. "Why do Chinese women practice contraception? the case of rural Jilin Province." *Studies in Family Planning* 22:39–51.

Den Bandt, M.L. 1980. "Voluntary childlessness in the Netherlands." *Journal of Family and Economic Issues* 3:329–349.

Dunson, D.B. and H. Zhou. 2000. "A Bayesian Model for Fecundability and Sterility." *Journal of the American Statistical Association* 95:1054–1062.

Elias, S. G., P. A. H. van Noord, P. H. M. Peeters, I. den Tonkelaar, and D. E. Grobbee. 2005. "Childhood Exposure to the 1944-1945 Dutch Famine and Subsequent Female Reproductive Function." *Human Reproduction* 20:2483–2488.

Farewell, VT. 1982. "The use of mixture models for the analysis of survival data with long-term survivors." *Biometrics* 38:1041–1046.

Gini, C. 1924. "Premières recherches sur la fécondabilité de la femme." In *Proceedings of the International Mathematics Congress*, 2, pp. 889–892.

Gluckman, P. and M. Hanson. 2006. *Developmental origins of health and disease*. Cambridge: Cambridge University Press.

Heckman, J.J. and J.R. Walker. 1990. "Estimating fecundability from data on waiting times to first conception." *Journal of the American Statistical Association* 85:283–294.

Henry, L. 1964. "Mortalité intra-utérine et fécondabilité." *Population (French Edition)* 19:899–940.

Hong, Y. 2006. "Marital decision-making and the timing of first birth in rural China before the 1990s." *Population studies* 60:329.

John, A.M., J.A. Menken, and A.K.M.A. Chowdhury. 1987. "The effects of breastfeeding and nutrition on fecundability in rural Bangladesh: A hazards-model analysis." *Population Studies* 41:433–446.

Jowett, A. J. 1991. "The Demographic Responses to Famine: The Case of China 1958-61." *GeoJournal* 23:135–146.

Kallan, J. and J.R. Udry. 1986. "The determinants of effective fecundability based on the first birth interval." *Demography* 23:53–66.

Kung, J.K. and J.Y. Lin. 2003. "The Causes of China's Great Leap Famine, 1959-1961*." *Economic Development and Cultural Change* 52:51–73.

Lambert, P.C. 2007. "Modeling of the cure fraction in survival studies." *Stata Journal* 7:351–375.

Lambert, PC, PW Dickman, CL Weston, and JR Thompson. 2009. "Estimating the cure fraction in population-based cancer studies by using finite mixture models." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 59:1–21.

Larsen, U. and J. Menken. 1989. "Measuring sterility from incomplete birth histories." *Demography* 26:185–201.

Larsen, U. and J.W. Vaupel. 1993. "Hutterite fecundability by age and parity: strategies for frailty modeling of event histories." *Demography* 30:81–102.

Lavely, W.R. 1986. "Age patterns of Chinese marital fertility, 1950-1981." *Demography* 23:419–434.

Leridon, H. 1981. "Fertility and contraception in 12 developed countries." *International Family Planning Perspectives* 7:70–78.

Leridon, H. 2008. "A new estimate of permanent sterility by age: Sterility defined as the inability to conceive." *Population Studies* 62:15–24.

Lin, J.Y. and D.T. Yang. 2000. "Food availability, entitlements and the Chinese famine of 1959-61." *The Economic Journal* 110:136–158.

Lucas, A. 1991. "Programming by early nutrition in man." In *The Childhood Environment and Adult Disease: Symposium, London, 15-17 May 1990*, edited by G. R. Bock and Whelan J., pp. 38–55. UK: John Wiley & Sons.

Lumey, L. H. and A. D. Stein. 1997. "In utero exposure to famine and subsequent fertility: The Dutch Famine Birth Cohort Study." *American Journal of Public Health* 87:1962–1966.

Lumey, L. H. and A. D. Stein. 2009. "Letter to the editor: Increased reproductive success of women after prenatal undernutrition?" *Human Reproduction* 00:1–1.

Lummaa, V. 2003. "Early Developmental Conditions and Reproductive Success in Humans: Downstream Effects of Prenatal Famine, Birthweight, and Timing of Birth." *American Journal of Human Biology* 15:370–379.

Maller, R.A. and X. Zhou. 1996. *Survival analysis with long-term survivors*. Chichester/New York: Wiley New York.

Menken, J., J. Trussell, and S. Watkins. 1981. "The nutrition fertility link: an evaluation of the evidence." *Journal of Interdisciplinary History* 11:425–441.

Miall, C.E. 1985. "Stigma of Involuntary Childlessness, The." *Social Problems* 33:268.

Moors, H. G. 1978. "The Netherlands Survey on Fertility and Parenthood Motivation, 1975: A Summary of Findings." In *Word Fertility Survey, No. 12*. International Statistical Institute.

Painter, R. C., R. G. J. Westendorp, S. R. de Rooij, C. Osmond, D. J. P. Barker, and T. J. Roseboom. 2008. "Increased reproductive success of women after prenatal undernutrition." *Human Reproduction* 23:2591–2595.

Peng, X. 1987. "Demographic consequences of the Great Leap Forward in China's provinces." *Population and Development Review* 13:639–670.

Rasmussen, K. M. 2001. "The "Fetal Origins" Hypothesis: Challenges and Opportunities for Maternal and Child Nutrition." *Annual Review of Nutrition* 21:73–95.

Rindfuss, R.R. and S.P. Morgan. 1983. "Marriage, sex, and the first birth interval: The quiet revolution in Asia." *Population and Development Review* 9:259–278.

Short, S.E., M. Linmao, and Y. Wentao. 2000. "Birth planning and sterilization in China." *Population Studies* 54:279–291.

Song, Shige. 2008. "Mortality Consequences of the 1959-1961 Great Leap Forward Famine in China: Debilitation, Selection, and Mortality Crossovers." California Center for Population Research On-Line Working Paper Series.

Song, Shige. 2009. "Does Famine Have a Long-Term Effect on Cohort Mortality? Evidence from the 1959-1961 Great Leap Forward Famine in China." *Journal of Biosocial Science* 41:469–491.

Song, Shige, Wei Wang, and Peifeng Hu. 2009. "Famine, Death, and Madness: Schizophrenia in Early Adulthood after Prenatal Exposure to the Chinese Great Leap Forward Famine." *Social Science & Medicine* 68:1315–1321.

Sposto, R. 2002. "Cure model analysis in cancer: an application to data from the Children's Cancer Group." *Statistics in medicine* 21:293–312.

Stearns, S.C. 1992. *The evolution of life histories*. Oxford University Press Oxford.

Stein, Z. 1975. *Famine and Human Development: The Dutch Hunger Winter of 1944-1945*. New York: Oxford University Press.

Stein, Z. and M. Susser. 1975. "Fertility, fecundity, famine: food rations in the dutch famine 1944/5 have a causal relation to fertility, and probably to fecundity." *Human Biology* 47:131–154.

Trussell, J. and C. Wilson. 1985. "Sterility in a population with natural fertility." *Population Studies* 39:269–286.

Walker, K.R. 1984. *Food grain procurement and consumption in China*. New York: Cambridge University Press.

Wang, F. and Q. Yang. 1996. "Age at marriage and the first birth interval: The emerging change in sexual behavior among young couples in China." *Population and Development Review* 22:299–320.

Weinstein, M. and M. Stark. 1994. "Behavioral and biological determinants of fecundability." *Annals of the New York Academy of Sciences-Paper Edition* 709:128–144.

Weinstein, M., J.W. Wood, M.A. Stoto, and D.D. Greenfield. 1990. "Components of age-specific fecundability." *Population Studies* 44:447–467.

Whiteford, L.M. and L. Gonzalez. 1995. "Stigma: the hidden burden of infertility." *Social Science & Medicine* 40:27–36.

Wood, J.W. 1994. *Dynamics of human reproduction: biology, biometry, demography*. New York: Aldine.

Wood, J.W., D.J. Holman, A.I. Yashin, R.J. Peterson, M. Weinstein, and M.C. Chang. 1994. "A multistate model of fecundability and sterility." *Demography* 31:403–426.

Yao, S. 1999. "A Note on the Causal Factors of China's Famine in 1959-1961." *Journal of Political Economy* 107:1365–1369.

Table 1: Urban-Rural Difference in Fertility Preference and Contraceptive Use

	Rural	Urban
Want Children	0.99	0.97
Don't Care	0.01	0.01
Don't Want Children	0.00	0.03
Contraceptive Use Before First Birth	0.01	0.06
Contraceptive Use After First Birth Or No Use	0.98	0.91
Timing Cannot Be Determined	0.01	0.03
Observations	7608	2953

Table 2: Descriptive Statistics of the Analytical Sample: Rural Residents with No Contraceptive Use Before First Birth and Want Children

	Pre-Famine Cohort	Famine Cohort	Post-Famine Cohort
No Education	0.311	0.203	0.151
Primary School	0.282	0.268	0.273
Junior High School	0.237	0.273	0.377
Senior High and Above	0.169	0.256	0.198
Ethnic Majority	0.924	0.918	0.915
Ever Pregnant	0.991	0.987	0.992
Ever Had Live Birth	0.989	0.985	0.990
Observations	3417	2450	4694

Table 3: Results from Mixture Logistic-Lognormal Model on Female Sterility and Fecundability in China: Time to First Pregnancy, N = 7,361

	Model 1	Model 2	Model 3
Sterility Model (Odds Ratios)			
Pre-Famine Cohort	0.50*	0.50*	0.50*
	[0.26,0.95]	[0.26,0.95]	[0.27,0.95]
Post-Famine Cohort	0.38**	0.40**	0.38**
	[0.20,0.71]	[0.21,0.75]	[0.20,0.71]
Primary School		0.85	
		[0.44,1.63]	
Junior High School		0.74	
		[0.36,1.53]	
Senior High and Above		1.11	
		[0.43,2.85]	
Ethnic Majority		0.87	
		[0.36,2.10]	
Fecundability Model (Time Ratios)			
Pre-Famine Cohort	1.04	1.02	1.02
	[0.99,1.09]	[0.98,1.07]	[0.98,1.07]
Post-Famine Cohort	0.93**	0.94**	0.94**
	[0.90,0.98]	[0.90,0.99]	[0.90,0.99]
Primary School		0.92***	0.92***
		[0.88,0.96]	[0.88,0.96]
Junior High School		0.88***	0.88***
		[0.84,0.92]	[0.84,0.92]
Senior High and Above		0.87***	0.87***
		[0.81,0.93]	[0.81,0.93]
Ethnic Majority		0.89***	0.89***
		[0.84,0.94]	[0.84,0.94]
AIC	54845.0	54807.3	54800.4
BIC	54893.4	54910.9	54876.3

Exponentiated coefficients; 95% confidence intervals in brackets

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Results from Mixture Logistic-Lognormal Model on Female Sterility and Fecundability in China: Time to First Birth, N = 7,354

	Model 1	Model 2	Model 3
Sterility Model (Odds Ratios)			
Pre-Famine Cohort	0.46*	0.45*	0.46*
	[0.25,0.85]	[0.24,0.84]	[0.25,0.85]
Post-Famine Cohort	0.38**	0.40**	0.38**
	[0.21,0.69]	[0.22,0.73]	[0.21,0.69]
Primary School		0.82	
		[0.45,1.52]	
Junior High School		0.66	
		[0.33,1.32]	
Senior High and Above		0.97	
		[0.39,2.45]	
Ethnic Majority		0.92	
		[0.39,2.14]	
Fecundability Model (Time Ratios)			
Pre-Famine Cohort	1.04**	1.03*	1.03*
	[1.01,1.08]	[1.00,1.07]	[1.00,1.07]
Post-Famine Cohort	0.98	0.98	0.98
	[0.95,1.00]	[0.95,1.01]	[0.95,1.01]
Primary School		0.93***	0.93***
		[0.91,0.96]	[0.91,0.96]
Junior High School		0.92***	0.92***
		[0.89,0.95]	[0.89,0.95]
Senior High and Above		0.91***	0.91***
		[0.87,0.96]	[0.87,0.96]
Ethnic Majority		0.90***	0.90***
		[0.86,0.93]	[0.86,0.93]
AIC	50283.2	50229.7	50223.3
BIC	50331.5	50333.3	50299.3

Exponentiated coefficients; 95% confidence intervals in brackets

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Results from Mixture Logistic-Lognormal Model on Female Sterility and Fecundability in China: Urban Population

	Pregnancy Model 1	Pregnancy Model 2	Birth Model 1	Birth Model 2
Sterility Model (Odds Ratios)				
Pre-Famine Cohort	1.39		1.39	
	[0.39,5.00]		[0.54,3.57]	
Post-Famine Cohort	1.75		1.67	
	[0.53,5.73]		[0.69,4.02]	
Fecundability Model (Time Ratios)				
Pre-Famine Cohort	1.00		1.00	
	[0.92,1.10]		[0.95,1.05]	
Post-Famine Cohort	0.96		0.98	
	[0.88,1.04]		[0.94,1.04]	
Primary School	0.82*	0.81*	0.91	0.91
	[0.67,1.00]	[0.66,0.99]	[0.80,1.02]	[0.80,1.02]
Junior High School	0.80*	0.79*	0.86**	0.85**
	[0.67,0.96]	[0.66,0.95]	[0.77,0.96]	[0.76,0.95]
Senior High and Above	0.83*	0.82*	0.90	0.90
	[0.69,0.99]	[0.69,0.98]	[0.81,1.01]	[0.81,1.00]
Ethnic Majority	0.78***	0.78***	0.88**	0.88**
	[0.67,0.89]	[0.68,0.89]	[0.81,0.96]	[0.81,0.96]
AIC	20040.7	20035.5	18263.4	18257.4
BIC	20105.3	20076.6	18328.0	18298.5

Exponentiated coefficients; 95% confidence intervals in brackets
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

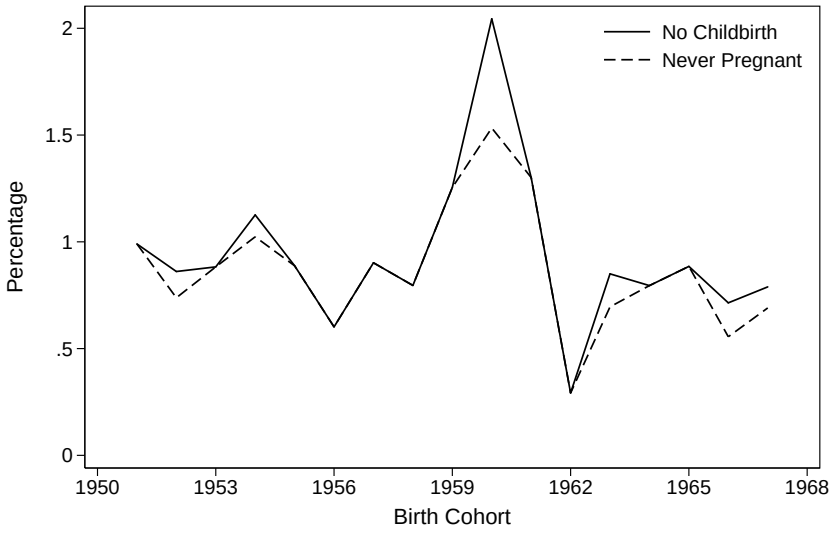

Figure 1: Cohort Trend in Infecundity, Rural Chinese Population

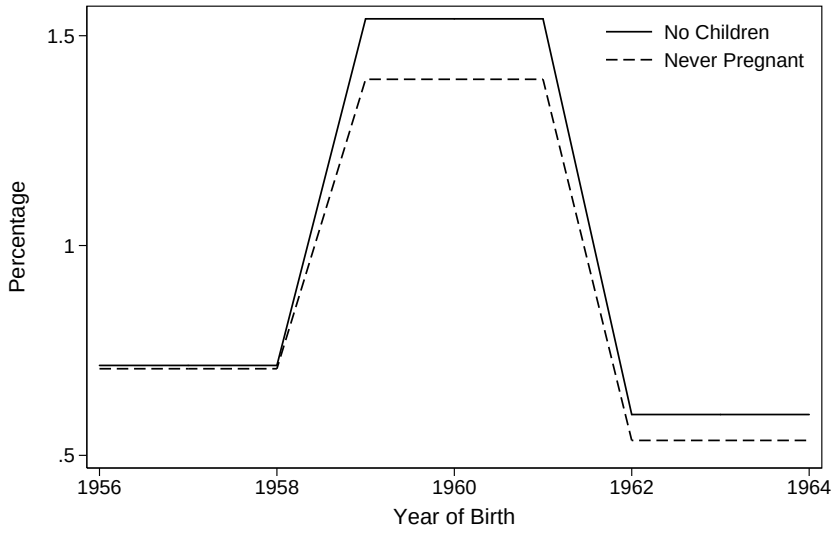

Figure 2: Predicted Cohort Trend in Sterility