DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

*DEMOGRAPHIC RESEARCH*

**VOLUME 29, ARTICLE 26, PAGES 707-728**
**PUBLISHED 8 OCTOBER 2013**
http://www.demographic-research.org/Volumes/Vol29/26/
DOI: 10.4054/DemRes.2013.29.26

*Research Article*

# Prenatal malnutrition and subsequent foetal loss risk: Evidence from the 1959-1961 Chinese famine

**Shige Song**

# Table of Contents

*Demographic Research*: Volume 29, Article 26
*Research Article*

# Prenatal malnutrition and subsequent foetal loss risk: Evidence from the 1959-1961 Chinese famine

**Shige Song**[1]

## Abstract

### BACKGROUND

Scientists disagree on whether prenatal malnutrition has long-term influences on women's reproductive function, and empirical evidence of such long-term effects remains limited and inconsistent.

### METHODS

Using the retrospective pregnancy history of 12,567 Chinese women collected in a nationally representative sample survey in 2001, this study conducted difference-in-differences analyses to investigate the relationship between prenatal exposure to the 1959-1961 Great Leap Forward Famine in China and the subsequent risk of involuntary foetal loss, including miscarriage and stillbirth, and how this relationship changes between the rural and urban populations.

### RESULTS

Prenatal exposure to the Great Leap Forward Famine had no long-term effect on women's risk of miscarriage. Such an exposure increased the risk of stillbirth among urban women but not among rural women.

### CONCLUSIONS

The results support the foetal origins hypothesis. The significant urban-rural difference in the effect of prenatal famine exposure on stillbirth suggests the presence of a long-term negative foetal origins effect and a strong selection effect caused by famine-induced population attrition.

[1] Queens College and CUNY Institute for Demographic Research, The City University of New York.
E-mail: shige.song@qc.cuny.edu.

## 1. Introduction

Scientists disagree on whether prenatal malnutrition has long-term influences on subsequent reproductive outcomes. On the one hand, the 'foetal origins' hypothesis and its straightforward extensions predict that prenatal malnutrition has a long-term negative influence on the female reproductive system by affecting the *in utero* development of the female foetus' organs that are responsible for the production and regulation of reproductive hormones (Drake and Walker 2004; Gardner, Ozanne, and Sinclair 2009; Lumey and Stein 1997; Lummaa 2003). On the other hand, based on the evolutionary life-history regulation theory, if prenatal malnutrition is associated with impaired health later in life, as predicted by the foetal origins hypothesis, an enhanced reproductive function is likely to present as a life-history compensation (Bateson et al. 2004; Painter et al. 2008). Both arguments are based on sound scientific theories; therefore the final adjudication depends on which argument is better supported by empirical evidence. Given the rampant poverty and malnutrition and high prevalence of involuntary foetal loss in many developing countries, testing the relationship between prenatal malnutrition and subsequent involuntary foetal loss risk is not only of important scientific value but also of significant policy relevance.

Due to the legal and ethical restrictions on the use of human subjects in scientific research, empirical evidence on the relationship between prenatal malnutrition and subsequent reproductive function has mainly been derived from observational studies. A number of such studies showed that women's prenatal conditions, as measured by their birth weight, are negatively associated with their subsequent risk of foetal loss and preterm birth, and positively associated with babies' birth weight and chance of survival (Emanuel et al. 1992; Hackman et al. 1983; Klebanoff, Meirik, and Berendes 1989; Little 1987; Sanderson, Emanuel, and Holt 1995; Skjærven, Wilcox, and Magnus 1997). However, due to the weaknesses of observational design, the causal nature of these findings remains unclear. More specifically, the observed association between women's birth weight and their subsequent reproductive outcomes may be the result of the causal mechanism articulated by the foetal origins hypothesis or the result of inadequate control of the common determinants of both women's birth weight and their reproductive outcomes that cannot be directly observed (Clarke 2005; Murnane and Willett 2010).

Famine provides an 'experiment-like' opportunity to address this important question. Similar to a randomised experiment, the assignment of the famine 'treatment' condition (i.e., famine-induced malnutrition) depends on exogenous forces that are unrelated to individual-level processes. This similarity between randomised experiments and famines has important design implications and may be the main reason that researchers increasingly utilise famines-based design to identify the causal effects of acute malnutrition when randomised experiments cannot be conducted (Chen and Zhou 2007; Huang et al. 2010; Song 2012; St Clair et al. 2005). However, it is important to emphasise the key differences between a famine-based natural experiment and a randomised experiment. These differences pose additional threats to the validity of famine-based studies that randomised experiment-based studies do not typically face, which may explain the conflicting results on the relationship between prenatal and early-life famine exposure and subsequent reproductive outcomes reported by earlier famine-based studies (Lumey and Stein 1997; Painter et al. 2008).

First, in randomised experiments, the researchers have full control over both the intensity and duration of the treatment conditions, whereas researchers cannot control the intensity or duration of famines. If the famine intensity or duration varies among different regions or groups and such variations are not adequately measured and controlled for, the statistical results can be biased. Second, although sample attrition can sometimes cause problems in randomised experiments, it always causes serious problems in famine-based studies. Famine is typically associated with increased mortality and reduced fertility (Bongaarts and Cain 1981). If such famine-induced population change is not random (i.e., famine survivors are significantly different from those who never experienced famine because the famine ‘weeded out’ the frail individuals and only the strong and healthy ones survived), simple cohort comparison between the famine and non-famine cohorts, which is the most commonly used identification strategy in famine-based studies, can lead to biased results. More recent studies confirmed that famine-induced population attrition is likely to be non-random (Gorgens, Meng, and Vaithianathan 2011; Song 2010), which creates selection effects that may change and even reverse the estimated relationship between famine exposure and health outcomes (Song, Wang, and Hu 2009).

The current study bridges this knowledge gap. Using the 1959-1961 Chinese Great Leap Forward Famine (GLFF) as a natural experiment and a ‘difference-in-differences’ (DD) method (Angrist and Pischke 2009; Meyer 1995; Murnane and Willett 2010), this study investigated the relationship between Chinese women’s prenatal and early-life exposure to acute malnutrition and subsequent risk of involuntary foetal loss, including miscarriage and stillbirth. The results show that ‘complete’ prenatal famine exposure (i.e., those whose entire *in utero* period, from conception to birth, occurred during famine) significantly increases women’s subsequent risk of stillbirth in urban areas but not in rural areas. By contrast, prenatal famine exposure has no significant long-term effect on women’s risk of miscarriage. Knowing that the GLFF-induced malnutrition influenced both the urban and rural populations whereas the GLFF-induced excess population attrition was primarily a rural phenomenon (Dikotter 2010; Lin and Yang 2000; Peng 1987; Song, Wang, and Hu 2009), these results suggest that prenatal exposure to the GLFF has both a negative foetal origins effect that increases the risk of stillbirth and a selection effect that ‘decreases’ the risk of stillbirth.

## 2. Materials and methods

The Great Leap Forward Famine of 1959-1961 in China, which led to over 30 million excess deaths and 33 million fertility losses during the three-year period (Ashton et al. 1984; Lin and Yang 2000; Peng 1987; Yao 1999), has a number of unique features (Dikotter 2010; Kung and Lin 2003; Lin and Yang 2000; Peng 1987). First, it persisted for over three years in most parts of China, which is much longer than any other large-scale famine in modern human history. Second, there was a significant regional variation in the level of malnutrition. Third, there was a drastic urban-rural difference in the famine-induced excess mortality. Due to its unusually long duration it is possible to use the GLFF to capture effects that are triggered by an extensive period of cumulative famine exposure (Song 2012). The significant regional variations in the levels of malnutrition provide an opportunity to estimate the fine-grained dose-response relationship between prenatal and early-life malnutrition and adult health outcomes (Chen and Zhou 2007). The drastic urban-rural difference in famine-induced excess mortality during the GLFF was not the result of the urban-rural difference in food market conditions, as in most other famine situations. Rather, it was the result of the Soviet-style state socialist system that was designed to accelerate urbanisation and industrialisation, especially the growth of heavy industry, at the expense of agricultural development and the interests of peasants (Ash 2006; Lin and Yang 2000; Wu and Treiman 2004). Under this type of system, urban status serves as a safety buffer that shields its owners from the extreme nutrition deprivation and associated adverse conditions that lead to starvation, regardless of the local food market conditions.

Existing studies based on the GLFF either relied on simple cohort comparison (Gorgens, Meng, and Vaithianathan 2011; Li et al. 2011; Song 2009, 2010; St Clair et al. 2005) or DD analysis that utilised two sources of variation (Chen and Zhou 2007; Huang et al. 2010a; Huang et al. 2010b; Song 2013). The validity of the simple cohort comparison method rests on the assumption that, in the counterfactual absence of the famine influence, there is no cohort difference in the outcome of interest between the famine and non-famine cohorts. If a non-zero counterfactual cohort difference exists between the famine and non-famine cohorts, simple cohort comparison leads to biased results. If this is the case, the DD method, which rests on a weaker assumption that the counterfactual cohort difference between the famine and non-famine cohorts, zero or not, does not change between places that were under the famine influence and those that were not, can be used to partial out the influence of such non-zero counterfactual cohort difference.

Only one study, conducted by Song, Wang and Hu (2009), utilised all three sources (temporal, spatial, and institutional) of variation in the GLFF impact simultaneously to isolate the long-term health consequences of prenatal malnutrition. This study showed that famine severity, measured as the famine-induced increase in infant mortality, significantly increased the risk of schizophrenia among the urban famine cohort but not among the rural famine cohort. These two sets of DD results (i.e., separately for the urban and rural samples), combined with the *a priori* knowledge that the rural population suffered much higher famine-induced excess mortality than the urban population, suggest the presence of both a long-term negative foetal origins effect and a selection effect caused by the famine-induced differential excess mortality.

The results reported by Song, Wang and Hu (2009) were based on a regression model that included three-way interaction between birth cohort, regional famine severity index, and urban-rural status. A similar model was estimated in this study as follows:

$$\log\left(\frac{P_{ik}}{1-P_{ik}}\right) = \alpha + \beta_1 U_i + \beta_2 S_k + \beta_3 \sum_{i=1956}^{1965} C_i + \beta_4 U_i \times S_i + \beta_5 \sum_{1956}^{1965} C_i \times U_i + \beta_6 \sum_{i=1956}^{1965} C_i \times S_k + \beta_7 \sum_{i=1956}^{1965} C_i \times S_k \times U_i \quad (1)$$

In Equation (1), the probability of miscarriage or stillbirth for the $i^{th}$ woman living in the $k^{th}$ region, $P_{ik}$, is modelled as the function of her urban-rural status, $U_i$, year of conception, $C_i$, the famine severity of her region, $S_k$, and all of the two-way and three-way interactions involved. The model is estimated using a generalised estimating equation (GEE) method to account for the clustering effect induced by the fact that each woman may have experienced multiple childbirths (Zeger and Liang 1986). Separate models were estimated for miscarriage and stillbirth.

A convenient way to extract quantities of interest (i.e., the DD estimates) from estimated regression results is through statistical simulation. After estimating the logistic regression model described in Equation (1), one can draw $R$ random samples (e.g., $R$ = 1,000) from the sampling distribution of model parameters, which is multivariate normal with mean $\beta$, the estimated coefficients vector, and variance matrix $V(\beta)$, the estimated variance and covariance matrix for the estimated coefficients. Using these $R$ simulated coefficient vectors, one can then produce simulated predicted probabilities, differences in predicted probabilities, and the associated confidence intervals (King, Tomz, and Wittenberg 2000; Zelner 2009).

For example, let $\Delta\Delta P_{1959u}$ be the DD estimate of the famine effect on miscarriage for the 1959 urban cohort. Treating the 1964 cohort as the reference group, the DD estimate can be obtained through the following equation:

$$\Delta\Delta P_{1959u} = \left[\left(P_{U=1,C=1959,S=max} - P_{U=1,C=1959,S=min}\right) - \left(P_{U=1,C=1964,S=max} - P_{U=1,C=1964,S=min}\right)\right] \tag{2}$$

in which $P_{U=1,C=1959,S=max}$, $P_{U=1,C=1959,S=min}$, $P_{U=1,C=1964,S=max}$ and $P_{U=1,C=1964,S=min}$ denote the predicted probabilities of having a miscarriage for urban women who (1) were conceived in 1959 and born in places of the maximum level of famine severity, (2) were conceived in 1959 and born in places of the minimum level of famine severity, (3) were conceived in 1964 and born in places of the maximum level of famine severity, and (4) were conceived in 1964 and born in places of the minimum level of famine severity. One can get simulated DD effects for all other cohorts by replacing the predicted probabilities for the 1959 cohort, $P_{U=1,C=1959,S=max}$ and $P_{U=1,C=1959,S=min}$, with the predicted probabilities for other cohorts of choice. One can also get DD estimates for the rural cohorts in similar ways.

The data used in this study come from the 2001 National Family Planning and Reproductive Health Survey (NFPRHS), a nationally representative sample survey conducted by the State Family Planning Commission of China. The survey collected information from 39,586 women (29,512 rural and 10,074 urban) aged 15 to 49 residing in family households in all 31 provincial administrative units in China through a stratified multistage sampling technique. All selected women were asked to provide their complete pregnancy history, including the year and month of pregnancy termination and the outcome of each pregnancy. The current analysis focused on the 12,567 women who were conceived between 1956 and 1965 and their 25,243 pregnancies.

The NFPRHS data are known to be of high quality and good population coverage (Chen, Xie, and Liu 2007; Zhang and Zhao 2006). However, for the purpose of studying foetal loss, two different kinds of underreporting problem deserve further discussion. First, the strict one-child family planning policy in China may discourage people from reporting higher order births. However, given that neither miscarriage nor stillbirth contributes to population growth, which is what the one-child policy aims to regulate, women have little incentive to lie about their history of miscarriage or stillbirth. In addition, the NFPRHS implemented various methods to reduce the underreporting of pregnancy and birth. For example, the sampling plan of NFPRHS was designed to be representative only at the national level so that local family planning officials would not be held responsible if true fertility levels in their areas were higher than previously reported (Chen, Xie, and Liu 2007). The second kind of underreporting comes from the difficulty in detecting very early pregnancies because human pregnancies remain virtually undetectable during their first 7-10 days, and, in some parts of the world, many women may not know that they are pregnant until many weeks or even months after conception (Wood 1994). On the one hand, this underreporting makes it difficult to derive a sensible estimate of the overall prevalence of foetal loss from self-reported pregnancy history information. On the other hand, as long as the underreporting does not change between cohorts and across regions systematically, it has little impact on the DD estimate of the famine effect on foetal loss.

The dependent variables, miscarriage and stillbirth, were constructed from the retrospective pregnancy history roster and measured dichotomously. For miscarriage, 1 = miscarriage and 0 = live birth; for stillbirth, 1 = stillbirth and 0 = live birth. Miscarriage is defined as spontaneous pregnancy loss before a foetus is viable outside a mother's womb, whereas stillbirth refers to pregnancy loss after the foetus could have survived. In practice, the distinction between miscarriage and stillbirth is mainly based on gestational age at which the death of the foetus occurs. A miscarriage refers to an early foetal death that occurs before the end of the second trimester, whereas a stillbirth refers to a late foetal death that occurs in the third trimester. Previous studies suggest that miscarriage is mainly associated with genetic defects such as chromosomal abnormalities, whereas stillbirth is more sensitive to non-genetic environmental factors of both the uterine and external environment (Brown 2008; Garcia-Enguidanos et al. 2002; Hoyert 1996). Comparisons of the results on miscarriage and stillbirth can reveal important information regarding the underlying mechanisms through which prenatal malnutrition exerts long-term influences on health and well-being.

In the current study, temporary variation of the famine impact was measured as mothers' year of conception, which was derived from the recorded information of their month and year of birth (i.e., by subtracting nine months from the year and month of birth of the child). Urban-rural residence at the time of the interview was used to capture the effect of the institutionalised inequality in minimum food security between the urban and rural Chinese populations and its shocking mortality consequences. Some women may have changed their place of residence between birth and the time of the interview. However, due to the household registration system that has been in place since the mid-1950s, only a highly selected minority (e.g., college graduates) could have crossed the urban-rural boundary, and the move was nearly always from rural to urban areas (Wu and Treiman 2004). The presence of a small fraction of rural-born women in the urban sample may lead to an estimate of the famine effect that is lightly biased toward the rural pattern, which should be interpreted with caution.

A province-level famine severity index was constructed as follows:

$$S_k = \frac{N^*_{1959,k}+N^*_{1960,k}+N^*_{1961,k}}{N_{1959,k}+N_{1960,k}+N_{1961,k}} \qquad (3)$$

where $N_{1959,k}$, $N_{1960,k}$, and $N_{1961,k}$ refer to the observed cohort size of the three famine cohorts for the $k^{th}$ province, and $N^{*}_{1959,k}$ , $N^{*}_{1960,k}$, and $N^{*}_{1961,k}$ refer to the expected cohort size of the three famine cohorts under the counterfactual condition that the famine never occurred. These expected counterfactual cohort sizes were produced by (1) extracting the size of the 1951-1970 birth cohorts for all 30 provinces using the public use sample of the 1982 Chinese census data (Minnesota Population Center 2010), (2) recoding the observed size of the 1959-1961 cohorts into missing values, and (3) imputing these ‘missing values’ using the multiple imputation algorithm proposed by Honaker and King (2010).[2] These expected counterfactual cohort sizes are inherently unknown quantities, and the uncertainty involved in predicting such unknown quantities cannot be fully represented by a single set of imputed values. The best practice is to impute the missing values multiple times, conduct statistical analysis using each imputed data set, and then combine the results to reflect our ignorance about these counterfactual cohort sizes. According to Rubin (1987), the final point estimate $\bar{\beta}$ is the average of the $M$ imputed complete-data point estimates $\beta_m$:

$$\bar{\beta} = \frac{1}{M}\sum_{m=1}^{M}\beta_m \qquad (4)$$

The final standard error is:

$$S.E. = \sqrt{W + \left(1 + \frac{1}{M}\right)B} \qquad (5)$$

in which the within-imputation variance $W$ is obtained as the average of the imputed data variance estimates $V^m$:

$$W = \frac{1}{M}\sum_{m=1}^{M}V^m \qquad (6)$$

and the between-imputation variance $B$ is obtained as:

[2] As Honaker and King (2010) pointed out, existing multiple imputation methods do not work well with time-series cross-section (TSCS) data structure (i.e., those with $T$ time points for each of $N$ cross-sectional units, as in the current case). The solution they proposed includes a new imputation model that allows smooth time trends, shifts across cross-sectional units, and correlation over time and space and a new bootstrapping-based expectation maximization imputation algorithm that is particularly efficient in imputing missing values with TSCS data structure.

$$B = \frac{1}{M-1} \sum_{m=1}^{M} \left( \beta_m - \bar{\beta} \right)^2 \quad (7)$$

To minimise the famine influence in generating the expected counterfactual cohort size, the size of the cohorts that are immediately adjacent to the three famine cohorts were also excluded (i.e., they did not contribute information to the imputation process). Summary information of the multiply imputed famine severity index is reported in Appendix Table A1.

Compared to the famine severity measures used in previous studies, the new index has some important advantages. First, the new famine severity index is based on publicly available information. Second, due to the power of the multiple imputation statistical technique, the new index adequately accounts for the uncertainty involved in predicting counterfactual quantities (i.e., the size of the famine cohort without the famine influence) and possible measurement errors.

In sensitivity analyses a number of other variables, including years of education, age at marriage, age at pregnancy termination, and pregnancy order, were also controlled. The addition of these control variables had little influence on the key coefficients and did not change the substantive conclusions. The main reason that these variables were not included in the main analysis is that they themselves may be influenced by prenatal or early-life famine exposure (Chen and Zhou 2007; Mu and Zhang 2010), which can potentially lead to endogeneity bias.

All statistical estimation and simulation in this study were carried out using open source statistical software *R* (R Core Team 2012). Multiple imputation was conducted using user-contributed package Amelia (Honaker, King, and Blackwell 2011), and statistical simulation was carried out using user-contributed package Zelig (Imai, King, and Lau 2008).

## 3. Results

Table 1 reports the famine exposure status for the 1956-1965 conception cohorts, the number of women in each cohort, the number of pregnancies they had, and the prevalence of miscarriage and stillbirth. Both the 1956 and 1957 cohorts experienced early-life famine exposure but not prenatal famine exposure. The 1958 cohort experienced early-life famine exposure and some prenatal famine exposure at a late gestational age. The size of the 1958 cohort was much smaller than the two preceding cohorts, suggesting the fertility-inhibiting effect of the famine. Both the 1959 and 1960 cohorts experienced early-life and full prenatal famine exposure because their entire *in utero* period, from conception to birth, occurred during the famine. The sizes of these cohorts were comparable to that of the 1958 cohort but much smaller than all other cohorts examined. Because the 1959 and 1960 cohorts were small and had the same famine exposure status, the two cohorts were grouped together in the multivariate analysis to increase the power of the analysis. The 1961 cohort experienced prenatal famine exposure and some early-life exposure. It is possible that in a small number of regions where the famine continued until 1963 (e.g., Jilin, Sichuan), both the 1962 and 1963 cohorts experienced partial prenatal famine exposure (at an early gestational age) and maybe even some early-life famine exposure. However, the overall famine impact was much weaker on these two cohorts than on the earlier cohorts, as evidenced by the difference in sizes of these cohorts. Finally, the 1964 and 1965 cohorts did not experience any form of famine exposure. The cohort patterns for stillbirth suggested that the 1957, 1958, and 1959 famine cohorts had the highest stillbirth prevalence. However, these three cohorts represent three distinctive combinations of prenatal and early-life famine exposure status, which makes it difficult to attribute the observed cohort pattern to famine exposure. The cohort patterns for miscarriage were even more obscure, suggesting that simple cohort comparison may not be adequate for revealing the underlying dynamics that regulate the patterns of foetal loss.

**Table 1: Exposure to the 1959-1961 Great Leap Forward Famine and the prevalence of miscarriage and stillbirth among selected conception cohorts**

| Year of Conception | Prenatal Famine Exposure | Early-Life Famine Exposure | *N* of Mothers | *N* of Pregnancies | Miscarriage (%) | Stillbirth (%) |
|---|---|---|---|---|---|---|
| 1956 | No | Yes | 1,250 | 2,799 | 3.61 | 1.07 |
| 1957 | No | Yes | 1,116 | 2,393 | 3.26 | 1.25 |
| 1958 | Some | Yes | 825 | 1,783 | 4.66 | 1.35 |
| 1959 | Yes | Yes | 954 | 1,915 | 3.12 | 1.24 |
| 1960 | Yes | Yes | 751 | 1,514 | 4.66 | 0.90 |
| 1961 | Yes | Some | 1,201 | 2,488 | 2.88 | 1.08 |
| 1962 | Some | Some | 1,779 | 3,437 | 2.55 | 0.75 |
| 1963 | Some | No | 1,545 | 3,005 | 3.09 | 1.06 |
| 1964 | No | No | 1,508 | 2,839 | 3.12 | 0.73 |
| 1965 | No | No | 1,638 | 3,070 | 3.16 | 1.01 |

Table 2 reports the coefficients (log odds) of the GEE logistic regression models for miscarriage and stillbirth. The model includes cohort, urban-rural residence, province-level famine severity index, all two-way and three-way interactions among them, and ethnicity. None of the coefficients in the miscarriage model were statistically significant, suggesting that miscarriage is not influenced by famine exposure status. By contrast, two coefficients in the stillbirth model were statistically significant. The negative coefficient for ethnic majority indicated that women from the ethnic majority group had a significantly lower risk of stillbirth than women from the ethnic minority group. The statistically significant three-way interaction term between the 1959-60 cohort, famine severity index, and urban residence suggested that stillbirth was influenced by women's prenatal famine exposure status.[3]

**Table 2: Coefficients (log odds) and standard errors from GEE logistic regression of miscarriage and stillbirth among Chinese women using multiply imputed data (M = 20)**

| | Miscarriage | | Stillbirth | |
|---|---|---|---|---|
| **Cohort [a]** | **Coefficient** | **Standard error** | **Coefficient** | **Standard error** |
| 1956 | 0.10 | 0.92 | 0.33 | 0.90 |
| 1957 | -0.40 | 0.73 | 1.38 | 0.78 |
| 1958 | 0.11 | 0.78 | 0.39 | 0.88 |
| 1959-60 | 0.25 | 0.69 | 1.33 | 0.85 |
| 1961 | 0.25 | 1.00 | 0.42 | 1.02 |
| 1962 | 0.25 | 0.74 | 0.78 | 0.85 |
| 1963 | -0.37 | 0.86 | -0.26 | 0.88 |
| 1965 | 0.34 | 0.77 | 0.48 | 0.83 |
| Urban | 0.50 | 0.86 | 0.24 | 1.30 |
| Famine severity (logged) | -0.17 | 0.92 | 1.27 | 0.93 |
| Ethnic Majority | -0.16 | 0.13 | -0.48* | 0.19 |
| **Cohort × urban** | | | | |
| 1956 × urban | 0.19 | 1.16 | 0.18 | 1.54 |
| 1957 × urban | -1.48 | 1.38 | -0.53 | 1.57 |
| 1958 × urban | -1.48 | 1.45 | -1.19 | 2.25 |
| 1959-60 × urban | 0.12 | 1.04 | -2.04 | 1.59 |
| 1961 × urban | 1.37 | 1.29 | 0.18 | 1.94 |
| 1962 × urban | -0.17 | 1.32 | -0.20 | 1.64 |
| 1963 × urban | 0.54 | 1.38 | 0.74 | 2.62 |
| 1965 × urban | 0.20 | 1.16 | 0.08 | 1.99 |

[3] Sensitivity analysis that excluded college graduates to reduce the effect of rural-to-urban migration yielded highly comparable results (available upon request).

**Table 2: (Continued)**

| Cohort [a] | Miscarriage Coefficient | Miscarriage Standard error | Stillbirth Coefficient | Stillbirth Standard error |
|---|---|---|---|---|
| **Cohort × famine severity** | | | | |
| 1956 × famine severity | -0.05 | 1.26 | 0.18 | 1.39 |
| 1957 × famine severity | 1.44 | 1.20 | -1.62 | 1.27 |
| 1958 famine severity | 1.29 | 1.35 | 0.51 | 1.38 |
| 1959-60 famine severity | 0.02 | 1.26 | -1.62 | 1.44 |
| 1961 famine severity | -1.32 | 2.08 | -0.41 | 1.64 |
| 1962 × famine severity | -0.38 | 1.33 | -1.46 | 1.36 |
| 1963 × famine severity | 0.70 | 1.47 | 0.96 | 1.30 |
| 1965 × famine severity | 0.11 | 1.32 | -0.30 | 1.23 |
| Urban × famine severity | 0.79 | 1.67 | -2.88 | 1.85 |
| **Cohort × urban × famine severity** | | | | |
| 1956 × urban × famine severity | -0.96 | 2.07 | 1.89 | 1.98 |
| 1957 × urban × famine severity | 1.20 | 1.99 | 2.93 | 2.50 |
| 1958 × urban × famine severity | 0.70 | 2.71 | 5.29 | 3.62 |
| 1959-60 × urban × famine severity | -1.03 | 1.80 | 5.69* | 2.34 |
| 1961 × urban × famine severity | -3.34 | 2.55 | 3.67 | 3.42 |
| 1962 × urban × famine severity | -1.53 | 2.56 | 4.32 | 2.61 |
| 1963 × urban × famine severity | -1.17 | 2.22 | -0.10 | 6.10 |
| 1965 × urban × famine severity | -1.42 | 2.04 | 1.32 | 3.91 |
| **N** | 24,994 [b] | | 24,421 [c] | |

*Note*: [a] The reference category is the 1964 cohort.
[b] Stillborn cases were excluded.
[c] Miscarried cases were excluded.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3 reports the simulated probabilities of stillbirth for each combination of birth cohort, urban-rural status, and famine severity levels based on the results reported in Table 2 (similar simulation results for miscarriage are not presented because none of the effects are significant). As a continuous variable the effect of famine severity was evaluated at values one and three, the lowest and highest median famine severity levels (Appendix Table 1). The table is divided into two panels, from left to right, that report the simulated results for the urban sample and rural sample separately. For both the urban and rural samples three different types of intermediate results were reported, including the simulated probabilities of stillbirth at a low famine severity level for each cohort, the simulated probabilities of stillbirth at a high famine severity level for each cohort, and the simulated DD in the probabilities of stillbirth between the high and low famine severity levels and between the selected cohort and the reference birth cohort.

**Table 3: Simulation results of the long-term famine effect on stillbirth among Chinese women using results obtained from the GEE logistic regression and multiply imputed data**

| | Urban | | | Rural | | |
|---|---|---|---|---|---|---|
| | **Severity = 1** | **Severity = 3** | **DD Estimate** | **Severity = 1** | **Severity = 3** | **DD Estimate** |
| 1956 | 0.008 | 0.007 | 0.005 | 0.005 | 0.031 | 0.012 |
| | [0.002, 0.021] | [0.001, 0.020] | [-0.012, 0.035] | [0.001, 0.013] | [0.012, 0.070] | [-0.021, 0.052] |
| 1957 | 0.019 | 0.005 | -0.008 | 0.016 | 0.013 | -0.017 |
| | [0.003, 0.067] | [0.000, 0.021] | [-0.063, 0.031] | [0.006, 0.032] | [0.004, 0.029] | [-0.049, 0.008] |
| 1958 | 0.006 | 0.134 | 0.134 | 0.007 | 0.042 | 0.021 |
| | [0.000, 0.035] | [0.006, 0.544] | [-0.002, 0.542] | [0.002, 0.020] | [0.012, 0.101] | [-0.018, 0.083] |
| 1959-60 | 0.002 | 0.021 | **0.025** | 0.017 | 0.011 | -0.019 |
| | [0.000, 0.008] | [0.004, 0.073] | **[0.001, 0.083]** | [0.006, 0.036] | [0.003, 0.035] | [-0.052, 0.013] |
| 1961 | 0.011 | 0.046 | 0.041 | 0.008 | 0.019 | -0.003 |
| | [0.001, 0.045] | [0.002, 0.204] | [-0.021, 0.192] | [0.002, 0.022] | [0.005, 0.049] | [-0.036, 0.034] |
| 1962 | 0.013 | 0.016 | 0.009 | 0.009 | 0.007 | -0.016 |
| | [0.002, 0.044] | [0.001, 0.067] | [-0.033, 0.071] | [0.003, 0.027] | [0.002, 0.018] | [-0.045, 0.006] |
| 1963 | 0.050 | 0.052 | 0.008 | 0.003 | 0.041 | 0.024 |
| | [0.000, 0.465] | [0.000, 0.771] | [-0.417, 0.692] | [0.001, 0.008] | [0.020, 0.079] | [-0.006, 0.063] |
| 1964 | 0.007 | 0.001 | --- | 0.004 | 0.018 | --- |
| | [0.000, 0.035] | [0.000, 0.008] | | [0.001, 0.011] | [0.007, 0.041] | |
| 1965 | 0.018 | 0.009 | -0.003 | 0.007 | 0.018 | -0.003 |
| | [0.001, 0.087] | [0.000, 0.076] | [-0.082, 0.068] | [0.003, 0.016] | [0.007, 0.037] | [-0.029, 0.021] |

*Note*: $DD_{C=1956,1957,\ldots,1965} = (P_{C=1956,1957,\ldots,1965,severity=3} - P_{C=1956,1957,\ldots,1965,severity=1}) - (P_{1964,severity=3} - P_{1964,severity=1})$.

Based on the DD results reported in Table 3, full prenatal famine exposure, as experienced by the 1959-1960 cohorts, significantly increased the risk of stillbirth among urban women but not among rural women. In addition, urban women born in 1958 and 1961, which are immediately adjacent to 1959-1960, also had an increased risk of stillbirth, although the effects are not statistically significant. The fact that four consecutive cohorts of urban women who were born around the time of the GLFF, all of which experienced at least some prenatal famine exposure, suffered from an increased risk of stillbirth strongly suggests that such an increase is real and famine exposure does have a long-term negative effect on female reproductive function. These findings are similar to the results reported by Song, Wang and Hu (2009) with respect to the effect of prenatal famine exposure on schizophrenia.

## 4. Discussion

The main findings of this study, i.e., (1) famine exposure does not have a long-term influence on women's risk of miscarriage, and (2) complete prenatal famine exposure leads to a significantly higher risk of stillbirth among urban women but not among rural women, support the prediction of the foetal origins hypothesis regarding the relationship between prenatal malnutrition and subsequent reproductive outcomes. The contrast between miscarriage and stillbirth is informative regarding the underlying mechanisms of such foetal origins effects. Miscarriage is mainly associated with genetic defects such as chromosomal abnormalities, whereas stillbirth is more sensitive to non-genetic environmental factors of both the uterine and external environment (Brown 2008; Garcia-Enguidanos et al. 2002; Hoyert 1996). The findings of this study thus suggest that the long-term negative effect of famine exposure on female reproductive function is through the interaction between the genetic and environmental factors, i.e., prenatal malnutrition disrupts the *in utero* development of the female foetus' reproductive organs, especially the uterus, which leads to an impaired adult uterine environment that is less capable of supporting foetuses. Finally, the urban-rural difference in the famine severity effect, which is consistent with the results on schizophrenia reported by Song, Wang and Hu (2009), suggests the co-existence of a foetal origins effect and a selection effect. Among the rural population, which suffered severe famine-induced excess mortality (Lin and Yang 2000; Peng 1987), the selection effect outweighs the negative foetal origins effect, whereas among the urban population, which suffered much less famine-induced excess mortality, the negative foetal origins effect outweighs the selection effect. Given that the majority of previous famine-based studies did not explicitly consider the potential influence of a selection effect, the results reported in the current study are particularly important and informative.

The success of a famine-based natural experimental study depends on the characteristics of the famine and the identification strategy employed. The GLFF is unique in many respects, including its long duration, policy-induced significant regional variations in the severity of food shortage, and the drastic urban-rural difference in the famine-induced excess mortality caused by the institutionalised predatory rural food procurement policy. Some of the unique features of the GLFF and its potential for accumulating human-based evidence regarding the effect of acute malnutrition were only recently recognized. The results reported in this study suggest that the combination of the GLFF and the DD method creates new opportunities for a more solid causal understanding of the relationship between prenatal malnutrition and health and well-being in later life by explaining some of the puzzling inconsistencies in earlier famine-based studies. But it should be noted that the identification strategy adopted in this study relies on the unique institutional settings of the Chinese society in the late 1950s and early 1960s. It may be useful for other studies based on the GLFF but will likely not be helpful in other famine contexts.

A number of weaknesses of this study are worth mentioning. First, because the pregnancy history information was self-reported, very early foetal losses, which are typically not noticeable without specialised instruments, were not included. This weakness, however, is likely to influences the results regarding miscarriage but not stillbirth. In addition, as long as the underreporting does not change between cohorts and across places systematically, the DD method can further reduce its influence on the estimated long-term famine effect. Second, there are signs of inadequate sample size, especially for the stillbirth model. For example, the DD estimates of the famine effect on stillbirth for the 1958 and 1961 urban cohorts were of the same direction as that for the 1959-1960 cohort and quite large in magnitude. Power analysis shows that such lack of statistical significance is likely due to the relatively small size of these two cohorts. It is important for future research to adequately address these issues with a larger sample.

## References


Angrist, J.D. and Pischke, J.S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Ash, R. (2006). Squeezing the peasants: Grain extraction, food consumption and rural living standards in mao's china. *The China Quarterly* 188: 959–998. doi:10.1017/S0305741006000518.

Ashton, B., Hill, K., Piazza, A., and Zeitz, R. (1984). Famine in china, 1958-61. *Population and Development Review* 10(4): 613–645. doi: 10.2307/1973284.

Bateson, P., Barker, D., Clutton-Brock, T., Deb, D., D'Udine, B., Foley, R.A., Gluckman, P., Godfrey, K., Kirkwood, T., Lahr, M.M., McNamara, J., Metcalfe, N.B., Monaghan, P., Spencer, H.G., and Sultan, S.E. (2004). Developmental plasticity and human health. *Nature* 430(6998): 419–421. doi:10.1038/nature 02725.

Bongaarts, J. and Cain, M. (1981). Demographic response to famine. In: Cahill, K. (ed.). *Famine*. New York: Orbis: 44–59.

Brown, S. (2008). Miscarriage and its associations. *Semin Reprod Med* 26(5): 391–400. doi:10.1055/s-0028-1087105.

Chen, J., Xie, Z., and Liu, H. (2007). Son preference, use of maternal health care, and infant mortality in rural china, 1989-2000. *Population Studies* 61(2): 161–183. doi:10.1080/00324720701340194.

Chen, Y. and Zhou, L. (2007). The long-term health and economic consequences of the 1959-1961 famine in china. *Journal of Health Economics* 26(4): 659–681. doi:10.1016/j.jhealeco.2006.12.006.

Clarke, K.A. (2005). The phantom menace: Omitted variable bias in econometric research. *Conflict Management and Peace Science* 22(4): 341–352. doi:10.1080/07388940500339183.

Dikotter, F. (2010). Mao's great famine: The history of china's most devastating catastrophe, 1958–1962. New York: Walker & Company.

Drake, A.J. and Walker, B.R. (2004). The intergenerational effects of fetal programming: Non-genomic mechanisms for the inheritance of low birth weight and cardiovascular risk. *Journal of Endocrinology* 180(1): 1–16. doi:10.1677/joe.0.1800001.

Emanuel, I., Filakti, H., Alberman, E., and Evans, S.J.W. (1992). Intergenerational studies of human birthweight from the 1958 birth cohort. 1. Evidence for a multigenerational effect. *BJOG: An International Journal of Obstetrics & Gynaecology* 99(1): 67–74. doi:10.1111/j.1471-0528.1992.tb14396.x.

Garcia-Enguidanos, A., Calle, M., Valero, J., Luna, S., and Dominguez-Rojas, V. (2002). Risk factors in miscarriage: A review. *European Journal of Obstetrics & Gynecology and Reproductive Biology* 102(2): 111–119. doi:10.1016/S0301-2115(01)00613-3.

Gardner, D., Ozanne, S., and Sinclair, K. (2009). Effect of the early-life nutritional environment on fecundity and fertility of mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364(1534): 3419–3427. doi:10.1098/rstb.2009.0121.

Gorgens, T., Meng, X., and Vaithianathan, R. (2011). Stunting and selection effects of famine: A case study of the great chinese famine. *Journal of Development Economics* 97: 99–111. doi:10.1016/j.jdeveco.2010.12.005.

Hackman, E., Emanuel, I., van Belle, G., and Daling, J. (1983). Maternal birth weight and subsequent pregnancy outcome. *JAMA* 250(15): 2016–2019. doi:10.1001/jama.1983.03340150058027.

Honaker, J. and King, G. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science* 54(2): 561–581. doi:10.1111/j.1540-5907.2010.00447.x.

Honaker, J., King, G., and Blackwell, M. (2011). Amelia ii: A program for missing data. *Journal of Statistical Software* 45(7): 1–47.

Hoyert, D.L. (1996). Medical and life-style risk factors affecting fetal mortality, 1989-90. *Vital and health statistics. Series 20, Data from the National Vital Statistics System* 20(31): 1–32.

Huang, C., Li, Z., Venkat Narayan, K.M., Williamson, D.F., and Martorell, R. (2010a). Bigger babies born to women survivors of the 1959--1961 chinese famine: A puzzle due to survival selection? *Journal of Developmental Origins of Health and Disease* 1(6): 412–418. doi:10.1017/S2040174410000504.

Huang, C., Li, Z., Wang, M., and Martorell, R. (2010b). Early life exposure to the 1959-1961 chinese famine has long-term health consequences. *Journal of Nutrition* 140(10): 1874–1878. doi:10.3945/jn.110.121293.

Imai, K., King, G., and Lau, O. (2008). Toward a common framework for statistical analysis and development. *Journal of Computational and Graphical Statistics* 17(4): 892–913. doi:10.1198/106186008X384898.

King, G., Tomz, M., and Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science* 44: 347–361. doi:10.2307/2669316.

Klebanoff, M.A., Meirik, O., and Berendes, H.W. (1989). Second-generation consequences of small-for-dates birth. *Pediatrics* 84(2): 343–347.

Kung, J.K. and Lin, J.Y. (2003). The causes of china's great leap famine, 1959-1961. *Economic Development and Cultural Change* 52(1): 51–73. doi:10.1086/380584.

Li, Y., Jaddoe, V.W., Qi, L., He, Y., Wang, D., Lai, J., Zhang, J., Fu, P., Yang, X., and Hu, F.B. (2011). Exposure to the chinese famine in early life and the risk of metabolic syndrome in adulthood. *Diabetes care* 34(4): 1014–1018. doi:10.2337/dc10-2039.

Lin, J.Y. and Yang, D.T. (2000). Food availability, entitlements and the chinese famine of 1959-61. *The Economic Journal* 110(460): 136–158. doi:10.1111/1468-0297.00494.

Little, R.E. (1987). Mother's and father's birthweight as predictors of infant birthweight. *Paediatric and Perinatal Epidemiology* 1(1): 19–31. doi:10.1111/j.1365-3016.1987.tb00084.x.

Lumey, L.H. and Stein, A.D. (1997). In utero exposure to famine and subsequent fertility: The dutch famine birth cohort study. *American Journal of Public Health* 87(12): 1962–1966. doi:10.2105/AJPH.87.12.1962.

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(3): 370–379. doi:10.1002/ajhb.10155.

Meyer, B.D. (1995). Natural and quasi-experiments in economics. *Journal of Business & Economic Statistics* 13(2): 151–161. doi:10.2307/1392369.

Minnesota Population Center (2010). Integrated public use microdata series, international: Version 6.0 [machine-readable database]. University of Minnesota.

Mu, R. and Zhang, X. (2010). Why does the great chinese famine affect the male and female survivors differently? Mortality selection versus son preference. *Economics & Human Biology* 9(1): 92–105. doi:10.1016/j.ehb.2010.07.003.

Murnane, R.J. and Willett, J.B. (2010). Methods matter: Improving causal inference in educational and social science research. New York: Oxford University Press.

Painter, R.C., Westendorp, R.G.J., de Rooij, S.R., Osmond, C., Barker, D.J.P., and Roseboom, T.J. (2008). Increased reproductive success of women after prenatal undernutrition. *Human Reproduction* 23(11): 2591–2595. doi:10.1093/humrep/den274.

Peng, X. (1987). Demographic consequences of the great leap forward in china's provinces. *Population and Development Review* 13(4): 639–670. doi:10.2307/1973026.

R Core Team (2012). R: A language and environment for statistical computing. In: *R Foundation Statistical Computing*. Vienna, Austria.

Rubin, D. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons. doi:10.1002/9780470316696.

Sanderson, M., Emanuel, I., and Holt, V.L. (1995). The intergenerational relationship between mother's birthweight, infant birthweight and infant mortality in black and white mothers. *Paediatric and Perinatal Epidemiology* 9(4): 391–405. doi:10.1111/j.1365-3016.1995.tb00162.x.

Skjærven, R., Wilcox, A.J., and Magnus, P. (1997). Mothers' birth weight and survival of their offspring: Population based study. *Bmj* 314(7091): 1376. doi:10.1136/bmj.314.7091.1376.

Song, S. (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. doi:10.1017/S0021932009003332.

Song, S. (2010). Mortality consequences of the 1959–1961 great leap forward famine in china: Debilitation, selection, and mortality crossovers. *Social Science & Medicine* 71(3): 551–558. doi:10.1016/j.socscimed.2010.04.034.

Song, S. (2012). Does famine influence sex ratio at birth? Evidence from the 1959–1961 great leap forward famine in china. *Proceedings of the Royal Society B: Biological Sciences* 279(1739): 2883–2890. doi:10.1098/rspb.2012.0320.

Song, S. (2013). Assessing the impact of in utero exposure to famine on fecundity: Evidence from the 1959–61 famine in China. *Population Studies* 67(3).

Song, S., Wang, W., and Hu, P. (2009). Famine, death, and madness: Schizophrenia in early adulthood after prenatal exposure to the chinese great leap forward famine. *Social Science & Medicine* 68(7): 1315–1321. doi:10.1016/j.socscimed.2009.01.027.

St Clair, D., Xu, M., Wang, P., Yu, Y., Fang, Y., Zhang, F., Zheng, X., Gu, N., Feng, G., Sham, P., and He, L. (2005). Rates of adult schizophrenia following prenatal exposure to the chinese famine of 1959–1961. *Journal of American Medical Association* 294(5): 557. doi:10.1001/jama.294.5.557.

Wood, J.W. (1994). Dynamics of human reproduction: Biology, biometry, demography. New York: Aldine.

Wu, X. and Treiman, D.J. (2004). The household registration system and social stratification in china: 1955-1996. *Demography* 41: 363–384. doi:10.1353/dem.2004.0010.

Yao, S. (1999). A note on the causal factors of china's famine in 1959-1961. *Journal of Political Economy* 107(6): 1365–1369. doi:10.1086/250100.

Zeger, S.L. and Liang, K.Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 42: 121–130. doi:10.2307/2531248.

Zelner, B.A. (2009). Using simulation to interpret results from logit, probit, and other nonlinear models. *Strategic Management Journal* 30(12): 1335–1348. doi:10.1002/smj.783.

Zhang, G. and Zhao, Z. (2006). Reexamining china's fertility puzzle: Data collection and quality over the last two decades. *Population and Development Review* 32(2): 293–321. doi:10.1111/j.1728-4457.2006.00119.x.

# Appendix

**Table A1: Descriptive statistics of the multiply imputed province-level famine severity index (M=20)**

| Province | Mean | Median | Standard Deviation | Minimum Value | Maximum Value |
|---|---|---|---|---|---|
| Beijing | 1.16 | 1.17 | 0.13 | 0.90 | 1.40 |
| Tianjin | 1.28 | 1.30 | 0.12 | 0.99 | 1.52 |
| Hebei | 1.50 | 1.53 | 0.08 | 1.28 | 1.64 |
| Shanxi | 1.27 | 1.27 | 0.05 | 1.19 | 1.37 |
| Neimengu | 1.22 | 1.24 | 0.11 | 0.88 | 1.37 |
| Liaoning | 1.40 | 1.38 | 0.11 | 1.20 | 1.64 |
| Jilin | 1.26 | 1.28 | 0.07 | 1.10 | 1.37 |
| Helongjiang | 1.21 | 1.20 | 0.07 | 1.11 | 1.36 |
| Shanghai | 1.00 | 1.01 | 0.10 | 0.77 | 1.13 |
| Jiangsu | 1.73 | 1.71 | 0.12 | 1.54 | 2.06 |
| Zhejiang | 1.53 | 1.54 | 0.06 | 1.40 | 1.66 |
| Anhui | 2.99 | 3.02 | 0.24 | 2.45 | 3.51 |
| Fujian | 1.40 | 1.39 | 0.08 | 1.23 | 1.52 |
| Jiangxi | 1.40 | 1.38 | 0.06 | 1.31 | 1.51 |
| Shandong | 1.74 | 1.73 | 0.06 | 1.65 | 1.86 |
| Henan | 2.07 | 2.05 | 0.12 | 1.87 | 2.35 |
| Hubei | 1.62 | 1.61 | 0.11 | 1.38 | 1.79 |
| Hunan | 1.97 | 1.98 | 0.10 | 1.77 | 2.14 |
| Guangdong | 1.49 | 1.49 | 0.04 | 1.40 | 1.56 |
| Guangxi | 1.75 | 1.73 | 0.11 | 1.50 | 1.95 |
| Sichuan | 2.49 | 2.51 | 0.18 | 2.08 | 2.91 |
| Guizhou | 2.03 | 1.98 | 0.14 | 1.85 | 2.36 |
| Yunnan | 1.78 | 1.78 | 0.13 | 1.47 | 2.03 |
| Xizang | 1.07 | 1.02 | 0.17 | 0.77 | 1.40 |
| Shaanxi | 1.20 | 1.20 | 0.06 | 1.10 | 1.32 |
| Gansu | 1.96 | 1.98 | 0.10 | 1.70 | 2.09 |
| Qinghai | 1.90 | 1.86 | 0.25 | 1.38 | 2.32 |
| Ningxia | 1.78 | 1.80 | 0.27 | 1.10 | 2.27 |
| Xinjiang | 1.35 | 1.36 | 0.10 | 1.20 | 1.56 |