Identifying the Intergenerational Effects of Prenatal Exposure to Nutritional Deprivation on Infant Mortality: Using the 1959-1961 Chinese Great Leap Forward Famine as a Natural Experiment

Shige Song

Queens College & CUNY Institute for Demographic Research
City University of New York

March 1, 2011

ABSTRACT

Using data from the 2001 Chinese National Family Planning and Reproductive Health Survey, I studied the relationship between prenatal exposure to the 1959-1961 Great Leap Forward Famine and the risk of infant death of the next generation. The results show that, on one hand, prenatal exposure to mild malnutrition reduced children's risk of infant death; on the other hand, prenatal exposure to severe malnutrition increased children's risk of infant death. Such a findings provides the first human-based supportive evidence to the developmental origins of health and disease argument and demonstrates the crucial role played by famine severity in determining the relationship between the effect of developmental plasticity and the effect of developmental disruption, the two distinctive forms of the developmental origins effects.

INTRODUCTION

The study of infant mortality has been benefitted from two different theoretical frameworks: the so-called "social model" and the so-called "medical model" (Frisbie, 2005). While the former emphasizes the importance of socioeconomic variables in determining infant survival, the latter focuses on the direct pathological pathways that lead to infant deaths (Wise, 1993). Recent progresses in the search for developmental origins of human health and disease suggest the possibility that infant mortality, among other reproductive health issues, may be traced to adverse maternal prenatal conditions,[1] especially poor nutritional conditions (Lumey and Stein, 1997). Such a new "developmental model" of infant mortality, if supported by empirical evidence, may provide an important new perspective through which old issues can be re-examined and new and more interesting questions can be addressed, which may then lead to improved policy making as well as more effective interventions.

Using data from the 2001 Chinese National Family Planning and Reproductive Health Survey, a large nationally representative sample survey conducted by the State Family Planning Commission of China, I identify and estimate the intergenerational effects of women's prenatal exposure to the 1959-1961 Chinese Great Leap Forward Famine on the risk of infant mortality of their children. Using famine as a natural experiment, an imputation-based famine severity measure that measures famine-induced cohort fertility loss, and a "difference-in-difference" identification strategy makes it possible to estimate a dose-response relationship between the severity of the prenatal famine exposure and the risk of infant mortality of the next children. Such a dose-response relationship is a crucial first step toward an understanding of the potentially heterogeneous nature of the effects of prenatal exposure to malnutrition on reproductive outcomes including infant mortality. In addition, such a relationship can help adjudicating between competing explanations and identifying the underlying causal mechanisms. As the results show, on one hand, in places where the level

[1] The term "maternal prenatal conditions" refers to prenatal conditions of mothers, not infants.

of famine-induced malnutrition was modest, prenatal famine exposure significantly *decreased* the risk of infant mortality of the next generation; on the other hand, such a "protective" effect of prenatal famine exposure decreased rapidly as the level of famine-induced malnutrition increased, and in places where famine severity was high, prenatal famine exposure *increased* the risk of infant mortality of the next generation. Such a pattern is consistent with the predictions made by developmental origins of health and disease argument regarding the effects of developmental plasticity and developmental disruption, and how the final results depend on the severity of the prenatal nutritional insult (Gluckman and Hanson, 2006; Gluckman et al., 2005).

The remainder of this article proceeds as follows. I first briefly review the relevant literature that connects prenatal famine exposure to the risk of infant death of the next generation, with special emphasis devoted to the distinction between the effect of developmental plasticity and the effect of developmental disruptions. I then introduce the empirical context – the 1959-1961 Great Leap Forward Famine in China. I next describe the data, variables, and identification and estimation strategy. At last, I present statistical results and discuss their significance and implications.

LITERATURE REVIEW

From Fetal Origins to Developmental Origins

The idea that prenatal conditions may have long-term influences on people's health and disease has a long tradition. But it was Barker and Osmond (1986) who, for the first time, articulated the idea in a systematic and coherent way, in the form of the so-called "fetal origins" hypothesis with some empirical support. Borrowing strength from evolutionary biology and life history theory (Bateson et al., 2004; Stearns, 1992), more recent development of the fetal origins hypothesis, in the form of the "developmental origins of health and disease" framework, began to pay attention to the potentially heterogeneous nature of the effect of prenatal exposure to adverse conditions

(Gluckman et al., 2005).

Lumey and Stein (1997) was the first to apply the fetal origins hypothesis in the study of reproductive health. They argued that prenatal exposure to severe malnutrition may have a long-term negative impact on women's reproductive function by disrupting the normal *in utero* development of fetus's organs responsible for production and regulation of female reproductive hormones. They used the 1944-1945 Dutch Famine as a natural experiment to test their hypothesis. Among the reproductive health indicators they considered, the only one that showed a fetal origins effect was perinatal death: stillbirth and infant death within the first week.

Using the same data, Painter et al. (2008) reached a drastically different conclusion. According to Painter et al., prenatal exposure to the Dutch Famine led to an *enhanced* female reproductive function, as evidenced by more children, more twin birth, earlier onset of childbirth, and lower likelihood of childlessness. Painter et al. (2008) attributed such findings to the effect of developmental plasticity, a well known phenomenon in evolutionary biology and life history theory (Bateson et al., 2004; Stearns, 1992). The idea is that the two traits, fertility and body maintenance are mutually balanced; and since the famine-born women faced increased risk of various chronic diseases which would lead to a reduced longevity, it should not be a surprise to see an enhanced reproductivity as a compensation.

Recent progress in the developmental origins of health and disease hypothesis provides an opportunity to synthesize the two different types of effects mentioned by Lumey and Stein (1997) and Painter et al. (2008). According to this argument, prenatal exposure to malnutrition can end up having either negative or positive effects on reproductive function, depending on a number of factors such as timing (gestation age) of the insult, the type of health outcome of interest, the match/mismatch of the prenatal and postnatal environments, and more importantly, the severity of the nutritional insult (Gluckman and Hanson, 2006; Gluckman et al., 2005). To be more specific, while fetuses can respond effectively and adaptively to mild nutritional insults in ways that can improve their reproductive success (as suggested by Painter et al.), their ability to do so attenuates

as the nutritional insults gets more severe. At some point, when the insult becomes too severe for fetuses to handle, normal prenatal development will be disrupted, causing both immediate and long-term damages (as suggested by Painter and Stein).

The demographic significance of the new developmental origins framework lies in the fact that this framework leads to a much more complicated picture regarding the relationship between nutrition and human fecundity and fertility than that has been described by earlier studies Bongaarts (1980). The idea that mild malnutrition may *reduce* infant mortality under some circumstances (and thus increases fertility) represents a drastic deviation from the currently dominant view, on which many population policies were based. Since mild malnutrition is still prevalent in many parts of the world, and the very same areas also suffer from the problem of overpopulation, it is pertinent to re-visit the issues regarding population growth in these areas without the preconception that malnutrition reduces fecundity and fertility.

Famine as Natural Experiment and Selection Effect

It is not a coincidence that both Lumey and Stein (1997) and Painter et al. (2008), the only two existing human-based empirical studies on the relationship between prenatal exposure to malnutrition and reproductive function, used famine as a natural experiment. On one hand, due to the nature of the relationship of interest, randomized trials are not option, for both legal and ethical reasons. In addition, results from animals cannot be readily generalized to humans. On the other hand, it is difficult to draw valid causal inferences from non-experimental designs using birth weight and other markers of early development as proxies for prenatal nutritional status: low-birth-weight mothers tend to give birth to low-birth-weight babies because of both genetic and environmental factors, which cannot be easily separated in non-experimental settings (Lummaa, 2003). By contrast, since famine-induced nutritional shock is exogenous (not influenced by individual-level processes), its effect can be measured by straightforward comparison between the exposed and the unexposed

groups and between groups with different levels of exposure.

As part of the cost of the natural experiment approach (as opposed to a real experiment in controlled settings), researchers cannot control the duration, severity, and magnitude of the famine to achieve the best designed effect. Instead, the researcher has to choose the famine that best meets the research purposes. However, as famines do not happen often in the modern world, this creates practical difficulties for researchers and may explain the conflicting results regarding prenatal exposure to the 1944-1945 Dutch Famine and female reproductive function in Lumey and Stein (1997) and Painter et al. (2008). [2]

In the context of famine, selection effect caused by differential mortality and/or differential fertility always complicates the results (Song, 2010). The simple fact that famine caused excess mortality/fetal loss and that "frail" people tended to be more badly influenced when famine hit suggests that people who survived famine (including people who were born during famine) tended to be better endowed than "ordinary" people who never experienced famine, and, unfortunately, such difference in health endowment is difficult, if not impossible, to directly measure and control for in general population surveys. As a result, cohort comparison between the famine cohort and the non-famine cohort with respect to health and longevity may not be a fair comparison because the famine cohort is constituted by famine survivors who tended to be the strongest and healthiest members while the non-famine cohort is constituted by more heterogeneous members. One has to be careful in interpreting cohort comparison results in the context of famine to make sure that the estimated cohort difference is real despite the selection effect.

[2] One possible reason that neither studies included a famine severity measures might be that the 1944-1945 Dutch Famine did not have significant regional variations in the severity of the famine-induced malnutrition since the famine occurred in a relative small area. This makes it difficult to show the effects of developmental plasticity and developmental disruption at the same time.

THE CURRENT STUDY

The current study uses the 1959-1961 Chinese Great Leap Forward Famine as a natural experiment to estimate the effects of prenatal exposure to malnutrition on the risk of infant mortality of the next generation.

The 1959-1961 Chinese Great Leap Forward Famine

A number of studies detail the causes and the magnitude of the 1959-1961 Great Leap Forward Famine in China (Ashton et al., 1984; Kung and Lin, 2003; Peng, 1987). According to these studies, the famine began in early 1959 and ended in 1961 in most parts of the country [3] and caused about 30 million excess deaths and 33 million fetal losses, making it one of the most disastrous tragedies in human history. The famine was national in scale with significant regional variations in its impact. For example, based on confidential data sources, (Peng, 1987, Table 2) reported an excess crude death rate for selected provinces ranging from 2.75 (Shanxi) to 109.78 (Sichuan), and similar patterns were reported by Lin and Yang (2000) based on other (also confidential) data sources.

Such a drastic regional variation in famine severity makes the 1959-1961 Great Leap Forward Famine an ideal natural experiment to study the does-response relationship between prenatal exposure to malnutrition and health (including reproductive health) outcomes later in life. This is not always possible in other famine contexts, and has not been done before.

[3] The famine lasted to 1962 in a number of provinces, including the most severely influenced province, Sichuan (Peng, 1987; Song et al., 2009).

Determine Prenatal Famine Exposure Status

Earlier research trying to assess the long-term effects of famine used birth cohort as the main indicator of famine exposure status. But since the focus of the present study is on the effect of *in utero* instead of early childhood famine exposure, I use information on"time of conception", which can be derived from the original information on "time of birth", as the main indicator of famine exposure.

Table 1 summarizes prenatal famine exposure status for women conceived in different years and months. The pre-famine cohort includes women conceived in 1955-1957; the famine cohort includes women conceived in 1959-1960, between September and December in 1958, and between January and April in 1961; the post-famine cohort includes women conceived in 1963-1965. Women conceived in the first eight months of 1958, the last eight months of 1961, and all twelve months of 1962 were excluded from the analysis because their prenatal famine exposure status could not be determined.

Such a cohort classification tried to balance between two competing goals: making the contrast between the famine and the non-famine cohorts, especially between the famine and post-famine cohorts, as sharp as possible while trying to allocate enough cases for the famine cohort. Future research utilizing larger sample size data can adopt a finer cohort classification that separates the effect between no prenatal exposure (1955-1957), some exposure at late gestation age (1958), full exposure (1959-1960), some exposure at early gestation age (1961-1962), and no exposure (1963-1965).

Measure Regional Variations in Famine Severity

Since no direct measure of the severity of the famine-induced malnutrition is not available, two types of indirect measures are sometimes used in the past research. The first measures famine-

induced mortality (Chen and Zhou, 2007), while the second measures the famine-induced fertility loss (Huang et al., 2010). The main problem with mortality-based measure is that it is not based on publicly available information so it cannot be verified, replicated, or further improved. In addition, since there has been no official estimates of the famine-induced death tolls and different estimates vary greatly, it is not clear how the numbers provided by Lin and Yang (2000), which is currently the only available province-level mortality information around the famine period, fits into the picture. In contrast, the fertility-based famine severity measure utilized publicly available information (public use sample of Chinese population census), so it can be replicated, checked, and further improved by other researchers.

The currently used fertility-based famine severity was calculated as the difference between the size of the famine cohorts and adjacent non-famine cohorts divided by the size of the non-famine cohorts. The underlying assumption is that the size of the adjacent non-famine cohorts provide a reasonable approximation of what the size of the famine cohorts would have been, had there been no famine. This is a questionable assumption, given the fact that the size of the adjacent non-famine cohorts were also influenced by the famine in one way of the other (excess infant death for the pre-famine cohorts and compensatory childbirth for the post-famine cohorts). In addition, by including only a small number of cohorts in the calculation, such a measure will not be able to adequately reflect the influence of long-term population dynamics.

In the present research, I construct and use a new famine severity measure based on famine-induced fertility change. Instead of using the information on the observed size of the famine cohorts and adjacent non-famine cohorts, I use the observed and the imputed size of the famine cohort to calculate the famine severity index. The basic idea is very simple: I treat the unobserved expected size of the famine cohort under normal non-famine condition as missing data and impute it from a long time series of cohort sizes. The new famine severity measures has a number of important strengths. First of all, the new measure substantively intuitive: it measures the difference between the observed cohort size (with famine impact) and the expected cohort size (without

famine impact). Second, since the imputation was based on an arbitrarily long time series of cohort size (the length of the time series is only limited by data availability), the imputed famine cohort size can adequately reflect any effects of the long-term population dynamics. Last but not least, by excluding certain birth cohorts from the imputation (e.g. cohorts immediately follow the famine), it is possible to produce different set of expected famine cohort sizes and conduct sensitivity tests to assess the robustness of the new imputation-based famine severity measure.

RESEARCH DESIGN

Data and Variables

I use data from the 2001 National Family Planning and Reproductive Health Survey (NFPRHS), a nationally representative sample survey conducted in 2001 by the State Family Planning Commission of China. The survey used a stratified multistage clustered sample to collect information in all 31 provincial administrative units in China from 39,586 women (29,512 rural and 10,074 urban) aged 15 to 49 living in family households. 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 dependent variable, whether a newly born child died at infancy, was constructed from the retrospective pregnancy-history roster and coded so that 1 = Died at infancy while 0 = Survived infancy. The key independent variable of interest is women's prenatal famine exposure status, measured as the year and month of conception. This variable has three categories: women conceived between 1955-1957 were considered pre-famine cohort, women conceived between September 1958 and February 1961 were considered the famine cohort, and women conceived between 1963 and 1965 were considered the post-famine cohort. Another key independent variable, famine severity, will be constructed from other data sources and merged back into the main data for statistical

analysis, following procedures described in the next subsection.

I also controlled for place of residence, mother's year of schooling, mother's age at childbirth, mother's ethnicity, and infant's birth order. Place of residence is a binary variable that indicates where the respondent lived at the time of the interview, coded so that 0 = living in a rural area and 1 = living in an urban area. Even though it is possible that some women may have changed their place of residence between birth and time of the interview, due to the household registration system that was in place since the late 1950s, only a selected minority (e.g. college graduates) could have crossed the urban-rural divide and the flow was almost always from rural to urban areas (Wu and Treiman, 2004). Mother's age at childbirth is a continuous variable ranging from 14.67 to 44.92, with mean value of 25.40 and standard deviation of 3.63. Since mother's age at childbirth has been found to be an important determinant of the risk of infant death (Frisbie, 2005), it is important to control for this source of variations to get unbiased effect of cohort difference and famine severity. Mother's year of schooling is a continuous variable ranging from 0 to 16, with mean value of 6.03 and standard deviation of 4.19. Womens education is an important socioeconomic indicator and has been widely used in population and health studies in developing countries where other socioeconomic indicators such as household income or occupation are not reliable or unavailable (Desai and Alva, 1998). Ethnicity is a binary variable coded so that 1 = Han ethnic majority and 0 = non-Han ethnic minorities. The only child-level covariate is birth order, which is measured as a ordinal scale variable with 1 = first birth, 2 = second birth, and 3 = third and higher order birth.

Tibet represents a distinctive case due to the political situations at the time. I excluded it from the analysis.

Table 2 reports descriptive statistics for the selected sample.

Construct Famine Severity Measure through Imputation

Province-level famine severity measure is constructed from the 1% public use sample of the 1982 Chinese census, made available through Minnesota Population Center (2010). The first step is to extract information on the size of the i^{th} birth cohort in the k^{th} province, N_{ik}, from the census data, where $i = 1950, 1951, ..., 1970$. This yields 30 cross-section (province-level) time series on birth cohort size. The second step is to recode the observed cohort size of the three famine cohorts, $N_{1959,k}$, $N_{1960,k}$, and $N_{1961,k}$, into missing values. The third step is to impute these "missing values" using the multiple imputation algorithm for cross-section time series data proposed by Honaker and King (2010), to get the expected counterfactual cohort sizes $N^*_{1959,k}$, $N^*_{1960,k}$, and $N^*_{1961,k}$. In the last step, the famine severity index for the k^{th} province, S_k, is computed as the ratio between the sum of the expected famine cohort sizes and the sum of the observed famine cohort sizes:

$$S_k = \frac{N^*_{1959,k} + N^*_{1960,k} + N^*_{1961,k}}{N_{1959,k} + N_{1960,k} + N_{1961,k}} \tag{1}$$

To partial out the influence of the famine in generating the expected counterfactual famine cohort size, in the second step described above, I excluded the size of the cohorts that are immediately adjacent to the three famine cohorts (by coding them into missing values) so they do not contribute any information to the imputation of the counterfactual cohort size of the three famine cohorts.

The results of the above mentioned procedures are reported in Table 3.

Generalized Estimating Equation, Difference-in-Difference Method, and Statistical Simulation

The NFPRHS data is inherently multilevel in nature in the sense that each selected women may have multiple pregnancies/childbirths, all of which will be present in the data. Such a clustered

data structure violates an important assumption underlying generalized linear models (GLM) and can lead to biased results. Generalized estimating equation (GEE) is an extension to GLM that handles such clustered data structure by introducing a "working correlation" [4] for the observed responses (Zeger and Liang, 1986). Coefficients in GEE has the same intuitive population averaged or marginal effect interpretation as GLM. Another nice feature of GEE is that the point estimate of coefficients is robust to misspecification of the working correlation, although good working correlation can improve the efficiency of the model. Model selection for GEE can be based on *QIC*, a variation of *AIC* for quasi-likelihood models (Pan, 2001). *QIC* can be used for model comparison between different model specifications and between the same model with different working correlation assumptions.

Let C_i be the cohort of the i^{th} woman, where $C_i = 1$ if she was conceived before the famine, $C_i = 2$ if she was conceived during the famine, and $C_i = 3$ if she was conceived after the famine. Let S_k be a measure of famine severity of the k^{th} region (e.g., province). The probability of infant death of the j^{th} birth of the i^{th} woman can be modeled as a logistic regression with GEE:

$$\log\left(\frac{P_{ij}}{1-P_{ij}}\right) = \alpha + \beta_1 X_{ij} + \beta_{21} C_{i1} + \beta_{22} C_{i2} + \beta_3 S_k + \beta_{41} C_{i1} S_k + \beta_{42} C_{i2} S_k + \varepsilon \qquad (2)$$

in which X_{ij} represents the value of control variables (e.g., age at marriage, education, ethnicity, and birth order) for the j^{th} child of the i^{th} woman, C_{i1} denotes the pre-famine cohort, and C_{2i} denotes the famine cohort (and C_{3i}, the post-famine cohort, is the reference category). Similarly, $C_{i1}S_k$ denotes the interaction between the pre-famine cohort and famine severity measure, and $C_{i2}S_k$ denotes the interaction between the famine-cohort and famine severity measure. I tried a few working correlations and it turns out that the most commonly used "exchangeable" working

[4] Commonly applied working correlations include "independent", "exchangeable", "autoregressive", and "unstructured". Each working correlation represents a different a prior expectation about the within-cluster variation in the dependent variable. One can try all applicable working correlations and decide on *QIC*.

correlation provides a reasonable balance.

In the case of linear regression model, the coefficient for the interaction term between the famine severity and the dummy variables for cohort, β_{41} and β_{42} represents the "difference-in-difference" estimate of the effect of prenatal and early childhood exposure to malnutrition on the dependent variable, whose statistically significance can be directly tested. However, since the dependent variable in the present research is binary, and Equation (2) represents a nonlinear statistical model, the interpretation of such interaction effects becomes much more complicated. As Ai and Norton (2003) demonstrated, due to the nonlinear nature of logistic regression model, the underlying interaction effect between two independent variables cannot be equated to the marginal effect of the interaction terms between these two independent variables. Berry et al. (2010) provided a set of concrete numerical examples demonstrating that: (1) statistically significant interaction terms do not necessarily mean statistically significant interaction effect, (2) a lack of statistically significant interaction terms do not necessarily mean a lack of statistically significant interaction effect, and (3) the sign and the magnitude of the interaction terms do not always agree with the sign and the magnitude of the underlying interaction effects.

One way to interpret interaction effects in nonlinear models is through statistical simulation. After successfully estimated the statistical model as described in Equation (2), one can draw M random sample (e.g., $M = 1,000$) from the multivariate normal distribution with mean β, the estimated coefficients vector from the regression, and variance matrix $V(\beta)$, the estimated variance-covariance matrix for the estimated coefficients from the regression. With the M simulated coefficient vectors in hand, one can then produce simulated predicted probabilities, difference in simulated predicted probabilities, as well as the associated confidence intervals (King et al., 2000; Zelner, 2009). [5]

The first step is to test the statistical significance of the interaction effect between cohort and

[5] Alternative approaches to interpret interaction effects in nonlinear models include the analytical approach such as delta method and Bayesian approach through Markov Chain Monte Carlo simulation.

famine severity. Following the suggestion by Zelner (2009) and Berry et al. (2010), this can be done by simulating the second difference of four predicted probabilities based on the estimated parameters from Equation

$$\begin{aligned}\Delta\Delta P(y=1) = &\left[P(y=1 \mid S_k^{max}, C_i=2) - P(y=1 \mid S_k^{min}, C_i=2)\right] \\ &- \left[P(y=1 \mid S_k^{max}, C_i=3) - P(y=1 \mid S_k^{min}, C_i=3)\right]\end{aligned} \tag{3}$$

and check if its 95% confidence interval includes zero.

Similar strategy can be used to simulate other quantities of interest. For example, one can test the statistical significance of the effect of famine severity on the probability of infant death for each of the three cohorts by simulating the first difference of the following form

$$\Delta P_1(y=1) = P(y=1 \mid S_k^{max}, C_i) - P(y=1 \mid S_k^{min}, C_i) \tag{4}$$

where $i = 1,2,3$ and check if the 95% confidence intervals include zero.

Or one can test the statistical significance of the difference in the probability of infant death between the famine and the post-famine cohorts at each level of famine severity by simulating the first difference of the following form:

$$\Delta P_2(y=1) = P(y=1 \mid C_i=2, S_k) - P((y=1 \mid C_i=3, S_k) \tag{5}$$

for each value of S_k and check if the 95% confidence intervals include zero.

RESULTS

Table 4 reports results from three logistic regressions with GEE on infant mortality. Model 1 only includes cohort and famine severity index, Model 2 also adds the interaction between the two, and Model 3 further adds a number of control variables. Based on *QIC*, Model 3 fits the data best.

Based on Model 3, the best fitted model, mother's education decreased infant mortality; ethnic majority had a lower level of infant mortality than the ethnic minority; and urban residents had a

lower level of infant mortality, although the effect was not statistically significant. These results agree with other studies of infant mortality in China. As for the biological factors, childbirth at older ages reduced infant mortality, while no statistically significant difference in infant mortality was found between different birth order.

Intergenerational Effects of Prenatal Exposure to Famine-Induced Malnutrition on Infant Mortality

Figure 1 shows the simulated "difference-in-difference" effect as depicted in Equation (3), based on the best fitted model in Table 4. To be more specific, it shows cohort difference in the changes in the predicted probability of infant death caused by a change in the famine severity from the minimum to the maximum level. Based on the 95% confidence interval, there is no statistically significant difference in the famine severity effect on infant death between the pre-famine and the famine cohorts or between the pre-famine and the post-famine cohorts. However, there is a statistically significant difference in the effect of famine severity effect on infant death between the famine and the post-famine cohorts. To be more specific, famine severity has a significantly larger effect (mean = 0.029, 95% C.I. = 0.004, 0.062) on infant mortality among the famine cohort than among the post-famine cohort.

The above results make intuitive sense because the severity of the famine is supposed to have influence on the famine cohort but not on the post-famine cohort. The second step is to make sure this is the case. However, the same result can also be produced other reasons and it is important to conduct further investigation on the relationship between famine severity and infant mortality for each cohort.

Figure 2 presents the predicted probabilities of infant death by famine severity for each cohort. In both the pre-famine and the post-famine cohorts, the relationship between famine severity and the probability of infant death seems to be modest; while in the famine cohort, the probability of

infant death rises steadily with famine severity.

Figure 3 further tests the statistical significance of the effect of famine severity on infant mortality for each cohort by presenting the difference in the predicted probabilities of infant death between the minimum and the maximum famine severity levels. Since both the pre-famine and the post-famine cohorts include zero value in their 95% confidence intervals, famine severity did not have statistically significant impact on the probability of infant death in these two cohorts. In contrast, the famine cohort does not include zero value in its 95% confidence interval, indicating that famine severity has a statistically significant effect on the probability of infant death in this cohort. In summary, since famine severity has significant effect only on the famine cohort but not on the pre-famine or the post-famine cohort, the difference-in-difference effect identified by Figure 1 is what it is hypothesized to be: the effect of prenatal exposure to famine-induced malnutrition.

By plotting the difference in the predicted probability of infant death between the famine and the post-famine cohorts at each level of the famine severity index, Figure 4 takes another look at the relationship of interest from a different perspective. At the lower end of the famine severity index, the famine cohort showed a significantly lower probability of infant death than the post-famine cohort; but such a cohort difference in the probability of infant mortality diminishes as the famine severity increases; when the famine severity reaches 1.7, the cohort difference in predicted infant death probability is no longer statistical significant. As famine severity continued to increase, the famine cohort started to show a higher probability of infant death than the post-famine cohort, but this difference remained only borderline significant.

Selection Effect or Developmental Origins Effects

How can these results speak to the central research question? The difference-in-difference results, as shown in Figure 1, suggested that prenatal famine exposure did have intergenerational effect on the risk of infant mortality; more in-depth investigation, as shown in Figure 2 and 3, suggested

that such effect worked in the expected way (i.e. famine severity changed infant mortality in the famine cohort but not in the post-famine cohort). Figure 4 further demonstrated that the effect of prenatal exposure status varied significantly with the level of famine severity: it reduced the risk of infant mortality at the lower end of famine severity but increased risk of infant mortality at the higher end of famine severity. This is, according to the developmental origins of health and disease framework, is how the effect of developmental plasticity and developmental disruptions work.

Such a pattern is not consistent with the effect of selection by differential mortality/fertility. Had selection effect played a central role, Figure 4 would have been completely different: the cohort difference in infant mortality between the famine and post-famine cohort should increase with famine severity because the more severe the famine the more positively selected the famine cohort becomes. It is apparently not the case here.

CONCLUSION

Using the 1959-1961 Chinese Great Leap Forward Famine as a natural experiment, this study identifies two different kind of effects of prenatal exposure to famine-induced malnutrition on the risk if infant mortality of the next generation. On one hand, prenatal exposure to moderate level of malnutrition reduced the risk of infant death; on the other hand, prenatal exposure to high level of malnutrition increased the risk of infant death, although this latter effect is only borderline significant.

These results provide the first set of human-based empirical evidence showing the dynamic nature of the developmental origins effects and the crucial role played by the severity of the malnutrition in determining the sign and the magnitude of the eventual combined effect. These results suggest that the relationship between nutritional status and human fertility is much more complicated than that has been shown in past research. Since infant mortality is an important fertility inhibiting factor in developing countries where the level of infant mortality is still high, the fact

that prenatal exposure to mild malnutrition may lead to a reduced infant mortality means that, everything else equal, women who were born in chronic malnutrition tended to have *higher* fertility than their well nourished as well as severely malnourished counterparts, due to the self-adjusting mechanisms that have been encoded into our genes. In light of these new findings, it is time to take a second look at some old wisdoms such as chronic malnutrition only has a minor effect on fecundity and fertility. As the present research suggested, without including a measure of famine severity into the analysis, the differential effect of prenatal famine exposure on infant mortality would have cancelled out to a large extent between the severely affected populations and the less severely affected populations, and one can easily jump to the conclusion that prenatal exposure to malnutrition had only a minor effect on infant mortality.

The combination of a natural experiment design, an imputation-based famine severity measure, and a difference-in-difference identification strategy has yielded important new findings. The unique feature of the Chinese Great Leap Forward Famine, a combination of a nation-wide impact and a significant regional variations in severity, makes it an ideal test case. Although the study was based on the Chinese population, the results should be applicable to other human populations as well.

One limitation of the present study is that, due to the relatively small sample size of the data used for the analysis, famine exposure status, measured in the form of conception cohort, is relatively crude. Ideally, the following prenatal famine exposure status should be compared and tested: (1) born before the famine with no prenatal famine exposure, (2) born during the famine, with partial prenatal famine exposure at late gestation age, (3) born during the famine, with full prenatal famine exposure, (4) born after the famine, with partial prenatal famine exposure at early gestation age, and (5) born after the famine with no prenatal famine exposure. The present study focused on the comparison between (1), (3), and (5) while excluded (2) and (4). Such limitation should be addressed in future research.

REFERENCES

Ai, C. and E. C. Norton. 2003. "Interaction terms in logit and probit models." *Economics Letters* 80: 123–129.

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. and C. Osmond. 1986. "Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales." *The Lancet* 327: 1077–1081.

Bateson, P., D. Barker, T. Clutton-Brock, D. Deb, B. D'Udine, R.A. Foley, P. Gluckman, K. Godfrey, T. Kirkwood, M.M. Lahr, et al. 2004. "Developmental plasticity and human health." *Nature* 430: 419–421.

Berry, W.D., J.H.R. DeMeritt, and J. Esarey. 2010. "Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential?" *American Journal of Political Science* 54: 248–266.

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

Chen, Y. and L. Zhou. 2007. "The Long-Term Health and Economic Consequences of the 1959-1961 Famine in China." *Journal of Health Economics* 26: 659–681.

Desai, S. and S. Alva. 1998. "Maternal education and child health: Is there a strong causal relationship?" *Demography* 35: 71–81.

Frisbie, W. P. 2005. "Infant mortality." In *Handbook of Population*, edited by Dudley L. Poston and Michael Micklin, pp. 251–282. New York: Kluwer Academic/Plenum.

Gluckman, P. D. and M. A. Hanson. 2006. "The conceptual basis for the developmental origins of health and disease." In *Developmental Origins of Health and Disease*, edited by P.D. Gluckman and M.A. Hanson, pp. 33–50. Cambridge: Cambridge University Press.

Gluckman, P. D., M. A. Hanson, H. G. Spencer, and P. Bateson. 2005. "Environmental influences during development and their later consequences for health and disease: implications for the interpretation of empirical studies." *Proceedings of the Royal Society B: Biological Sciences* 272: 671.

Honaker, J. and G. King. 2010. "What to Do about Missing Values in Time-Series Cross-Section Data." *American Journal of Political Science* 54: 561–581.

Huang, C., Z. Li, M. Wang, and R. Martorell. 2010. "Early Life Exposure to the 1959-1961 Chinese Famine Has Long-Term Health Consequences." *Journal of Nutrition* 140: 1874.

King, G., M. Tomz, and J. Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44: 347–361.

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.

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

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.

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.

Minnesota Population Center. 2010. *Integrated Public Use Microdata Series, International: Version 6.0 [Machine-readable database]*. Minneapolis: University of Minnesota.

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.

Pan, W. 2001. "Akaike's information criterion in generalized estimating equations." *Biometrics* 57: 120–125.

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

Song, S. 2010. "Mortality Consequences of the 1959-1961 Great Leap Forward Famine in China: Debilitation, Selection, and Mortality Crossovers." *Social Science & Medicine* 71: 551–558.

Song, S., W. Wang, and P. 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.

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

Wise, P. H. 1993. "Confronting racial disparities in infant mortality: Reconciling science and politics." *American Journal of Preventive Medicine* 9: 7–16.

Wu, X. and D. J. Treiman. 2004. "The Household Registration System and Social Stratification in China: 1955-1996." *Demography* 41: 363–384.

Zeger, S. L. and K. Y. Liang. 1986. "Longitudinal data analysis for discrete and continuous outcomes." *Biometrics* 42: 121–130.

Zelner, B. A. 2009. "Using simulation to interpret results from logit, probit, and other nonlinear models." *Strategic Management Journal* 30: 1335–1348.

Table 1: Time of Conception and Prenatal Exposure to the 1959-1961 Great Leap Forward Famine

Time of Conception		Prenatal Famine Exposure	Cohort
Year	Month		
1955-1957	1-12	No Exposure	Pre-Famine Cohort
1958	1-8	No or Some Exposure	—
1958	9-12	Some or Full Exposure	Famine Cohort
1959-1960	1-12	Full Exposure	Famine Cohort
1961	1-4	Some or Full Exposure	Famine Cohort
1961	5-12	No or Some Exposure	—
1962	1-12	No or Some Exposure	—
1963-1965	1-12	No Exposure	Post-Famine Cohort

Table 2: Descriptive Statistics of the Analytical Sample

	Mother's Cohort		
	Pre-Famine	Famine	Post-Famine
Mother Information			
Ethnicity			
Ethnic Majority (%)	90.23	89.08	88.83
Ethnic Minority (%)	9.77	10.92	11.17
Place of Residence			
Urban (%)	17.20	21.07	16.45
Rural (%)	82.80	78.93	83.55
Years of Schooling	4.99	6.43	6.75
	(4.37)	(4.37)	(3.75)
Age at Childbirth	26.49	25.65	24.58
	(3.82)	(3.54)	(3.32)
Infant Information			
Birth Order			
First Birth (%)	46.03	50.66	53.50
Second Birth (%)	33.13	32.54	34.03
Third and Higher Order Birth (%)	20.84	16.80	12.47
Infant Death (%)	1.71	0.98	1.59
N of Mothers	3,397	1,921	4,606
N of Infants	7,380	3,792	8,610

Source: Author's calculations of the National Family Planning and Reproductive Health Survey, 2001.

Table 3: Observed and Imputed Cohort Size of the 1959-1961 Birth Cohorts and the Province-Level Famine Severity

Province	Cohort Size		Severity Index
	Observed	Imputed	
Beijing	5934	6796	1.145
Tianjin	4589	7916	1.725
Hebei	23558	36673	1.557
Shanxi	12532	18331	1.463
Neimenggu	10672	12681	1.188
Liaoning	19796	29775	1.504
Jilin	12525	16132	1.288
Heilongjiang	18454	21604	1.171
Shanghai	8013	7969	0.994
Jiangsu	23206	38555	1.661
Zhejiang	18063	27044	1.497
Anhui	9873	35284	3.574
Fujian	11380	16371	1.439
Jiangxi	13988	19488	1.393
Shandong	28156	50812	1.805
Henan	22861	50664	2.216
Hubei	18552	32521	1.753
Hunan	17438	37892	2.173
Guangdong	26141	36228	1.386
Guangxi	12730	22208	1.745
Sichuan	25491	72099	2.828
Guizhou	8082	15488	1.916
Yunnan	11099	18229	1.642
Xizang	957	1041	1.088
Shaanxi	14611	17570	1.203
Gansu	6441	13074	2.030
Qinghai	1306	2247	1.721
Ningxia	1338	2877	2.150
Xinjiang	5481	7925	1.446

Source: The 1% public use sample of the 1982 Chinese population census.

Table 4: Results from GEE Model for Infant Mortality

	Model 1	Model 2	Model 3
Cohort Difference:			
– Pre-Famine cohort	0.08	-0.14	-0.15
	(0.14)	(0.31)	(0.33)
– Famine Cohort	-0.50*	-1.50***	-1.51**
	(0.21)	(0.44)	(0.47)
– Post-Famine Cohort	—	—	—
Famine Severity Index	0.18	-0.18	-0.34
	(0.23)	(0.33)	(0.35)
Cohort and Famine Severity Interaction:			
– Pre-Famine cohort × Famine Severity		0.40	0.49
		(0.49)	(0.53)
– Famine Cohort × Famine Severity		1.80**	1.94**
		(0.68)	(0.74)
Years of Schooling			-0.05**
			(0.02)
Han Ethnic Majority			-0.79***
			(0.16)
Urban Residence			-0.30
			(0.21)
Age at Childbirth			-0.08**
			(0.03)
Birth Order:			
– First Birth			—
– Second Birth			-0.06
			(0.16)
– Third and Higher Order Birth			0.08
			(0.26)
Constant	-4.44***	-4.23***	-1.07
	(0.17)	(0.20)	(0.63)
Observations	19685	19685	19685

Standard errors in parentheses

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

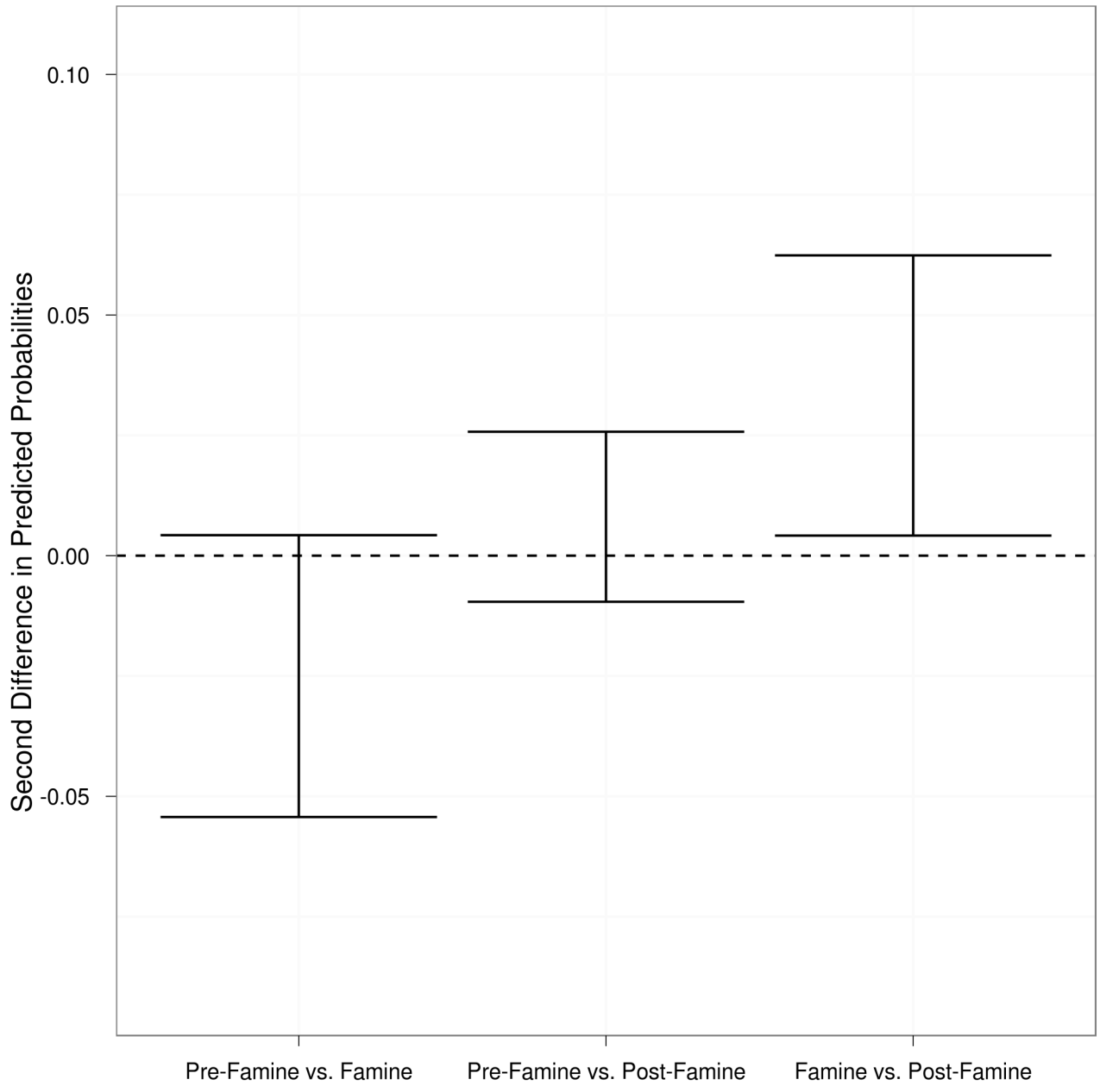

Figure 1: 95% Confidence Intervals of the Simulated Second Difference in Predicted Infant Death Probabilities between Different Cohorts and between the Highest and the Lowest Severity Levels

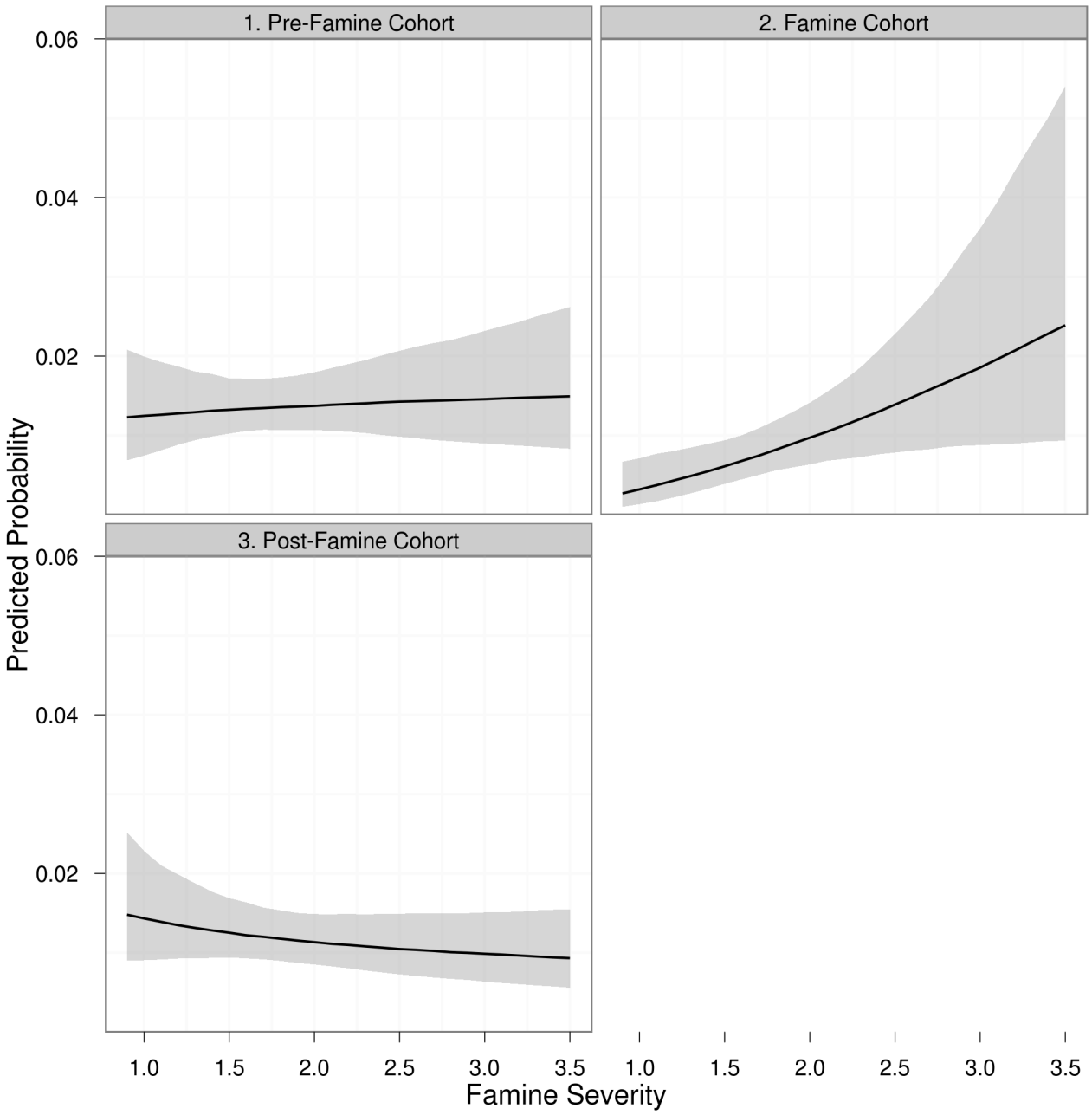

Figure 2: Predicted Probability (and 95% Confidence Intervals) of Infant Death by Mother's Birth Cohort and Famine Severity

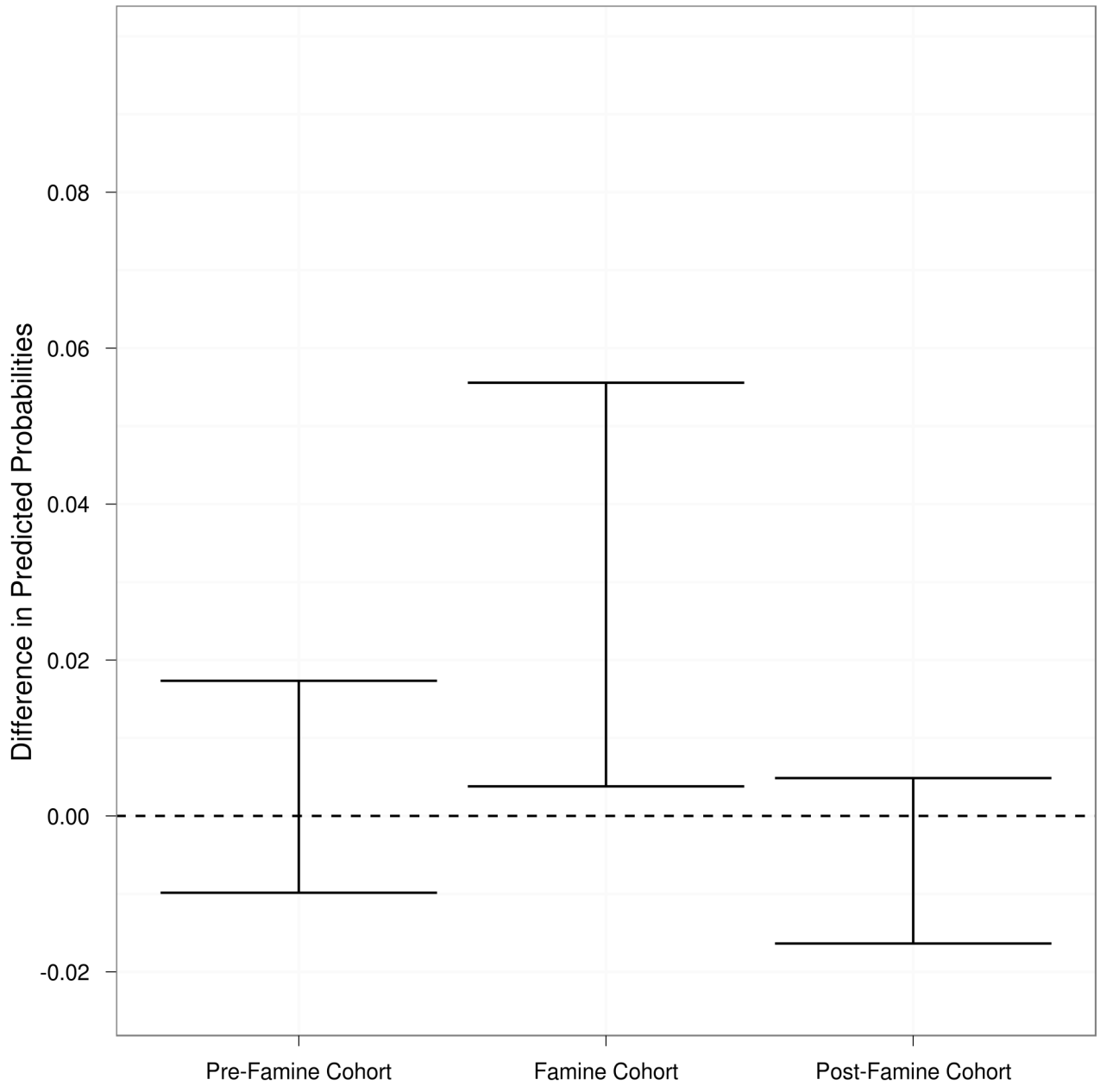

Figure 3: 95% Confidence Intervals of the Difference in Predicted Infant Death Probabilities between the Highest and the Lowest Famine Severity Levels for Each Cohort

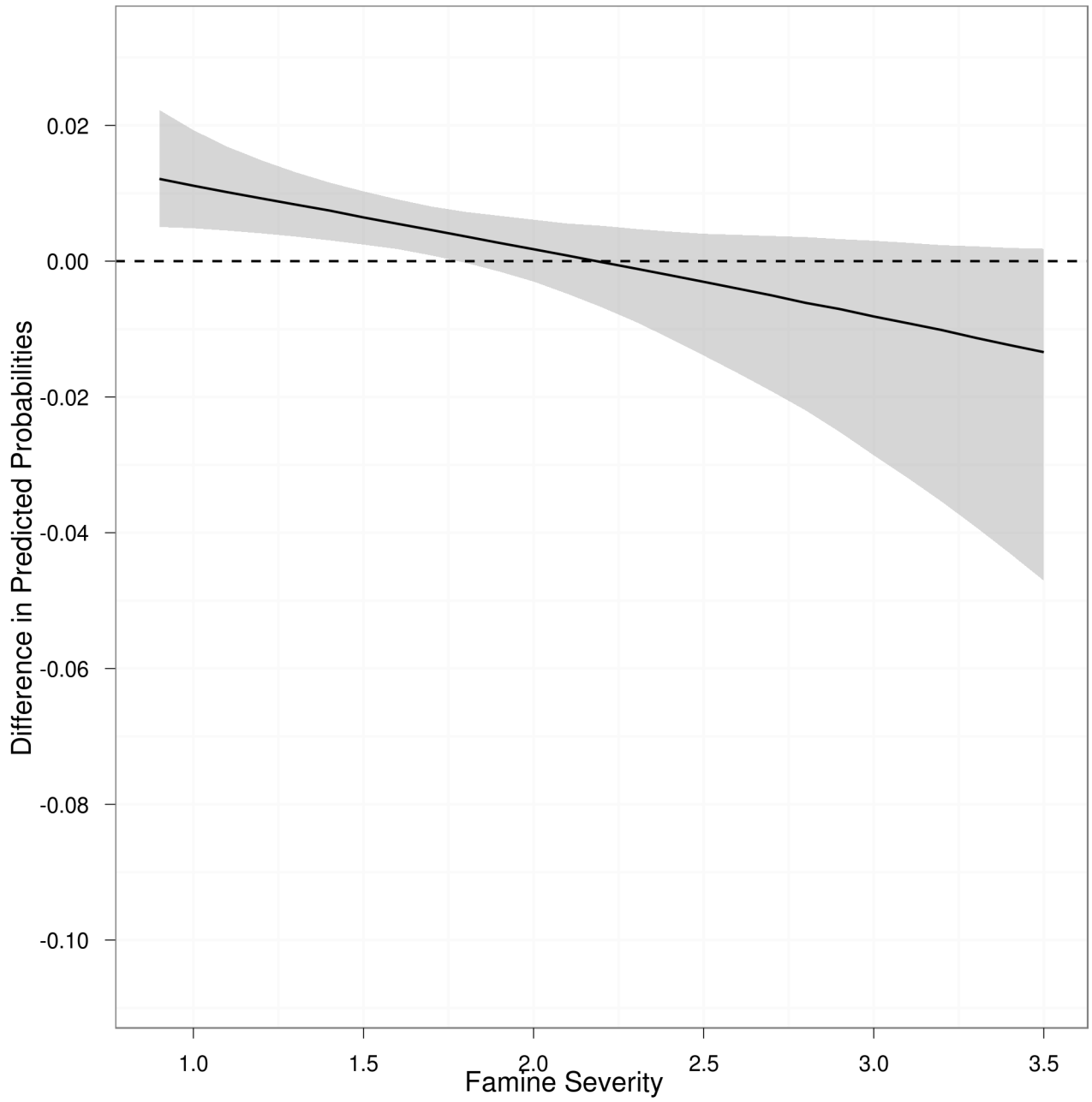

Figure 4: 95% Confidence Intervals of the Difference in Predicted Infant Death Probabilities between the Post-Famine Cohort and the Famine Cohort at the Minimum, the Median, and the Maximum Severity Levels