PAA 2012 manuscript No.
(will be inserted by the editor)

Identifying the Long-Term Effect of Prenatal Famine Exposure on Female Sterility (Extended Abstract)

Shige Song

April 2, 2012

1 Introduction

Female sterility, also known as "primary" and "permanent" female sterility, refers to the phenomenon that some women permanently lost their biological reproductive capacity even before reaching childbearing age. It is an important yet poorly understood aspect of human reproductive process.

This paper examines the possible long-term effect of prenatal exposure to the 1959-1961 Great Leap Forward Famine in China on female sterility. Famine-based natural experimental design has important advantages over observational design in tackling the endogenous selection issue and can lead to valid causal inference when conducting randomized experiment is not an option. Earlier natural experimental studies, which were all based on the 1944-1945 Dutch Famine, led to inconsistent results (Elias et al., 2005; Lumey and Stein, 1997; Painter et al., 2008). It is important for independent studies, and preferably based on other empirical settings, to replicate and verify these findings.

Compared to these earlier studies, the present study has a number of strengths. First of all, by combining two large nationally representative sample surveys, the current study utilizes a much larger study sample than all of the Dutch Famine-based studies. Second, unlike the Dutch Famine-based studies, which relied on simply cohort comparison to identify the long-term famine effect, the current study makes use of two distinctive sources of famine-related variations, famine exposure status and famine severity, to identify the long-term effect of famine on female sterility. Third, the use of multiple imputation-based famine severity index that captures the famine-induced cohort attrition provides an unique opportunity to obtain a valid baseline estimate of the famine effect on female sterility with little confounding of selection effect. Such a baseline estimate may play an important role in conceptually separating selection effect from the hypothesized biological effect of interest when selection effect is not ignitable.

2 Data and Method

The data sources used in this study come from two of the fertility surveys conducted by the Chinese State Family Commission in 1997 ($N = 15,000$ women) and 2001 ($N = 39,586$ women). The final analytical sample includes 9,969 women of childbearing age who were themselves conceived in 1956-1957 (the pre-famine cohort), 1959-1960 (the famine cohort), and 1962-1963 (the post-famine cohort).

2.1 Definition of Sterility

In 1997, the members of the youngest birth cohorts (i.e., the 1963 cohort) in the final analytical sample were 34 years old. To ensure a fair comparison between different cohorts and between different survey waves, I defined female sterility as not having a lively born baby up to age 34. Considered that most Chinese women got married by their mid 20s, this is a much more strict criterion than the commonly used "two years after marriage" or "seven years after marriage" (Bongaarts, 1980; Larsen, 2000; Liu et al., 2004), and thus lead to a lower yet more reasonable estimate of the sterility prevalence level.

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

2.2 Definition of Prenatal Famine Exposure

Table 1 shows the definition of prenatal famine exposure status. The post-famine cohort was completely unexposed to the famine and thus represents the reference group against which all other cohorts were compared against. The pre-famine cohort was exposed to the famine at early childhood but not prenatally, whereas the famine cohort experienced complete prenatal famine exposure (i.e., both at conception and at birth). Note that women who were conceived in 1958 and 1961 were excluded from the analysis because the presence of significant within cohort variation in famine exposure status. Such a classification minimizes the potential measurement error and ensures the validity of the cohort comparison results.

2.3 Famine Severity Index

Famine-induced cohort attrition also played an important role in the analysis. In order to calculate the famine-induced cohort attrition, one needs (1) the observed famine cohort size and (2) the expected famine cohort size under the counterfactual condition that the famine did not occur. Because the expected famine cohort size is intrinsically unobserved and unobservable, any imputations or predictions of such information is probabilistic and, if not controlled for properly, is subject to measurement error and underestimate of the variability involved.

Multiple imputation for missing data provides a reasonable solution to the problem (Rubin, 1987). Table 2 reports multiply imputed famine-induced cohort attrition ($M = 10$) using the imputation algorithm proposed by Honaker and King (2010). The imputed famine severity index was then merged into the main data set for statistical analysis. As the last step, the M different sets of estimated coefficients (point estimates and standard errors) were combined based on Rubin's rule of combination to construct the final results.

3 Results

3.1 Results from Simple Cohort Comparison

Model 1 and 2 in Table 3 estimate cohort difference (odds ratios) between the pre-famine, famine, and post-famine cohorts in female sterility risk. Both models show that the pre-famine cohort has slightly lower sterility risk than the post-famine cohort while the famine cohort has slightly higher sterility risk than the post-famine cohorts. Neither of the cohort differences is statistically significant.

Based on AIC, adding the famine severity index significantly improves the model fit. However, a statistically significant odds ratio of 0.40 does not make intuitive sense: why female sterility is lower in places where the famine was more severe? The answer to this question cannot be answered with these simple models.

3.2 Results from Difference-In-Difference Analysis

Model 3 and 4 in Table 3 estimate the difference-in-difference effect of famine exposure on female sterility risk. Based on AIC, both models fit the data significantly better than the simpler models (Models 1 and 2).

According to the best-fit model, in places where the famine-induced cohort attrition was negligible, the pre-famine cohort had lower sterility risk than the post-famine but the effect was not statistically significant, whereas the famine cohort had a statistically significantly higher sterility risk than the post-famine cohort. For the pre-famine cohort, the sterility risk increased with famine severity although such effect was not statistically significant. For the famine cohort, the sterility risk decreased as the famine severity increased and such effect was statistically significant.

Because logistic regression is nonlinear, the presence of the interaction terms between birth cohort and famine severity index complicates coefficient interpretation (Ai and Norton, 2003). One way to facilitate coefficient interpretation is to conduct statistical simulation based on the estimated coefficients and the variance-covariance matrix (King et al., 2000; Zelner, 2009).

Figure 1 shows the results from one such simulation. The simulated cohort difference in female sterility risk between the pre-famine and post-famine cohorts increase slightly with famine severity but, as the 95% confidence intervals show, such a cohort difference in female sterility risk was not statistically significant regardless of the level of famine severity. By contrast, the simulated cohort difference in female sterility risk between the famine and post-famine cohorts changes significantly with famine severity: (1) at the low end of the famine severity index, the famine cohort shows significantly higher female sterility risk than the post-famine cohort; (2) such a difference decreases as famine severity increases and eventually losses statistical significance when famine severity approaches 1.5 (i.e., approximately 33% of the original cohort was lost during the famine); (3) at the high end of the famine severity index, the famine cohort shows a lower female sterility risk than the post-famine cohort but such difference is not statistically significant.

3.3 Selection vs. Biological Effects

From previous studies on the Chinese famine, we know that even in places where famine did not cause massive excess mortality or fetal loss (e.g., in Beijing, Shanghai, Inner Mongolia, etc.), the negative impact of severe food shortage on people's daily life during the famine period was clear and significant. Given that the only source of selection effect is differential mortality and/or fertility reduction, the presence of a special group that was subject to the famine influence but did not experience massive famine-induced excess mortality or fertility reduction provides an opportunity to isolate the famine effect (which was "mild" when compared to the famine impact in other places where the famine caused much higher excess mortality or fertility loss) that is not confounded by severe selection effect.

Clearly, as Figure 1 shows, without selection effect, prenatal famine exposure leads to significantly increased female sterility risk (for the famine cohort), compared to the post-famine cohort. As famine severity gets stronger and stronger, more and more members of the original famine cohort was lost during the famine, selection effect begins to play an increasingly important role in shaping the observed cohort pattern in female sterility risk.

4 Conclusions

Combining two large national representative sample fertility survey, the present study identified a significant increase in female sterility risk among the famine cohort in places where the famine did not cause severe cohort attrition. Such a finding support the developmental origins of health and disease argument that, when facing adverse environmental conditions, the developmental plasticity mechanisms of human fetuses can lead to adaptive changes that can help them survive the adverse conditions; the same changes, however, may have long-term negative health and developmental consequences. The results also suggest the presence of a strong selection effect that dominates the cohort pattern in female sterility risk in places where the level of famine-induced cohort attrition was high. Unfortunately, there is no information in the data set that can be used to directly control for or estimate the selection effect. This is the subject that future research (with more suitable data sources) can make the most significant contribution to.

References

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

Bongaarts, J. 1980. “Malnutrition and fecundity.” *Studies in Family Planning* 11: 401–406.

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

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.

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.

Larsen, U. 2000. “Primary and secondary infertility in sub-Saharan Africa.” *International Journal of Epidemiology* 29: 285.

Liu, J., U. Larsen, and G. Wyshak. 2004. “Prevalence of primary infertility in China: in-depth analysis of infertility differentials in three minority province/autonomous regions.” *Journal of Biosocial Science* 37: 55–74.

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.

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.

Rubin, D.B. 1987. *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons.

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 the Status of Prenatal Exposure to the 1959-1961 Great Leap Forward Famine

Time of Conception		Prenatal Famine Exposure	Cohort
Year	Month		
1956-1957	1-12	No Exposure	Pre-Famine Cohort
1958	1-12	No or Some Exposure	—
1959-1960	1-12	Full Exposure	Famine Cohort
1961	1-12	No or Some Exposure	—
1962-1963	1-12	No Exposure	Post-Famine Cohort

Table 2 Multiply Imputed Province-Level Famine Severity Index ($M=10$)

	Multiply Imputed Values										Mean
	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$	$m=8$	$m=9$	$m=10$	
Beijing	1.06	1.32	1.06	1.15	1.24	1.16	1.24	1.06	1.27	1.34	1.19
Tianjin	1.51	1.40	1.34	1.44	1.49	1.47	1.38	1.65	1.35	1.12	1.42
Hebei	1.58	1.63	1.48	1.53	1.59	1.65	1.5	1.49	1.58	1.67	1.57
Shanxi	1.38	1.35	1.31	1.46	1.22	1.27	1.38	1.34	1.35	1.31	1.34
Neimenggu	1.28	1.35	1.12	1.29	1.12	1.14	1.29	1.28	1.32	1.25	1.24
Liaoning	1.55	1.37	1.53	1.40	1.45	1.38	1.50	1.49	1.44	1.37	1.45
Jilin	1.39	1.19	1.20	1.32	1.33	1.19	1.12	1.28	1.25	1.35	1.26
Heilongjiang	1.27	1.18	1.25	1.25	1.19	1.26	1.34	1.24	1.37	1.23	1.26
Shanghai	1.26	1.06	1.09	1.03	1.25	1.15	1.09	1.02	1.09	1.34	1.14
Jiangsu	1.83	1.56	1.68	1.80	1.73	1.77	1.75	1.86	1.87	1.47	1.73
Zhejiang	1.54	1.46	1.64	1.61	1.46	1.49	1.63	1.48	1.59	1.57	1.55
Anhui	2.86	3.13	3.06	3.16	2.90	3.28	3.34	3.18	2.84	3.07	3.08
Fujian	1.37	1.33	1.75	1.44	1.55	1.41	1.60	1.46	1.79	1.57	1.53
Jiangxi	1.51	1.47	1.40	1.44	1.43	1.36	1.37	1.40	1.51	1.35	1.42
Shandong	1.75	1.69	1.69	1.84	1.87	2.27	2.02	1.78	1.95	1.96	1.88
Henan	2.20	2.28	2.13	2.13	2.18	2.04	2.08	2.26	1.99	2.05	2.13
Hubei	1.61	1.76	1.84	1.66	1.75	1.68	1.79	1.78	1.60	1.67	1.71
Hunan	2.06	2.20	2.16	2.08	2.11	2.01	2.18	2.22	2.17	2.19	2.14
Guangdong	1.56	1.53	1.59	1.53	1.61	1.49	1.58	1.54	1.48	1.47	1.54
Guangxi	1.88	1.90	1.72	1.82	1.75	1.97	1.53	1.72	1.74	1.83	1.79
Sichuan	2.56	2.40	2.56	2.53	2.58	2.44	2.48	2.41	2.52	2.65	2.51
Guizhou	2.21	2.09	1.89	2.03	2.52	2.00	1.76	1.95	2.08	1.88	2.04
Yunnan	1.74	1.88	2.09	1.75	1.58	1.83	1.72	1.81	1.83	1.84	1.81
Xizang	1.18	1.25	.946	1.21	1.11	1.60	1.48	1.20	1.08	.993	1.21
Shaanxi	1.31	1.32	1.30	1.33	1.32	1.29	1.48	1.33	1.22	1.51	1.34
Gansu	2.15	2.16	2.1	2.03	2.05	1.93	2.02	2.00	1.78	1.7	1.99
Qinghai	1.82	1.72	1.79	2.2	2.05	2.25	1.62	1.97	1.58	1.93	1.89
Ningxia	1.84	2.02	2.09	1.75	1.92	1.65	1.82	1.84	1.33	1.92	1.82
Xinjiang	1.25	1.30	1.41	1.46	1.28	1.33	1.42	1.52	1.21	1.41	1.36

Source: The 1% public use sample of the 1982 Chinese Population Census.

Table 3 Results from Simple Logistic Regression Models Using Multiply Imputed Data, N=9,969

	Model 1	Model 2	Model 3	Model 4
Birth Cohort				
– Pre-Famine Cohort	0.82	0.81	0.67	0.70
	[0.57,1.20]	[0.56,1.17]	[0.26,1.70]	[0.28,1.76]
– Famine Cohort	1.33	1.25	3.33**	3.22**
	[0.93,1.89]	[0.87,1.79]	[1.41,7.85]	[1.37,7.57]
– Post-Famine Cohort	1.00	1.00	1.00	1.00
	[1.00,1.00]	[1.00,1.00]	[1.00,1.00]	[1.00,1.00]
Famine Severity Index (Log-Trasformed)		0.40**	0.61	0.70
		[0.20,0.79]	[0.23,1.61]	[0.25,1.97]
Birth Cohort and Famine Severity Interaction				
– Pre-Famine Cohort × Famine Severity			1.46	1.57
			[0.29,7.28]	[0.32,7.77]
– Famine Cohort × Famine Severity			0.11*	0.12*
			[0.02,0.65]	[0.02,0.71]
– Post-Famine Cohort × Famine Severity			1.00	1.00
			[1.00,1.00]	[1.00,1.00]
Han Ethnic Majority				0.80
				[0.48,1.33]
Post-Famine Conditions				1.01
				[0.92,1.11]
Years of Schooling				1.05**
				[1.01,1.09]
Observations	9969	9969	9969	9969
AIC	1754.31	1749.08	1744.89	1742.83

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

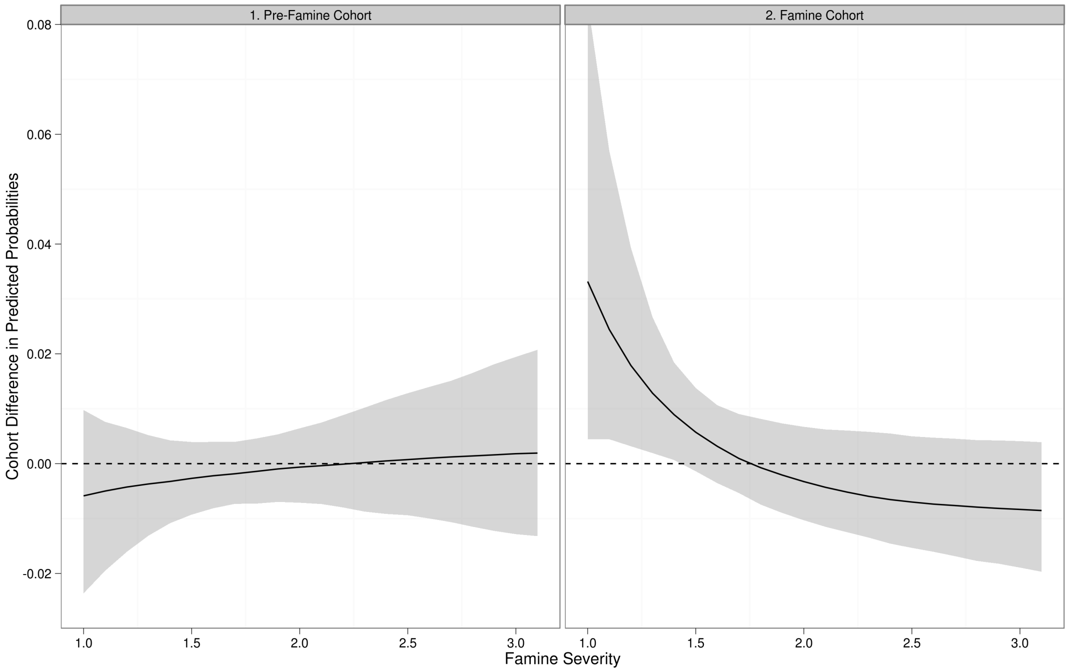

Fig. 1 Simulated cohort difference (between selected cohorts and the post-famine cohort) in the predicted sterility risk at each level of famine severity index