Social Science & Medicine 148 (2016) 110–122

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Is natural experiment a cure? Re-examining the long-term health effects of China's 1959–1961 famine

Hongwei Xu [a, *], Lydia Li [b], Zhenmei Zhang [c], Jinyu Liu [d]

[a] Institute for Social Research, University of Michigan, 426 Thompson St, ISR 2459, Ann Arbor, MI 48104-2321, USA
[b] School of Social Work, University of Michigan, USA
[c] Department of Sociology, Michigan State University, USA
[d] School of Social Work, Columbia University, USA

ARTICLE INFO

Article history:
Received 13 July 2015
Received in revised form
20 October 2015
Accepted 19 November 2015
Available online 2 December 2015

Keywords:
China
Fetal origins hypothesis
Famine
Biomarker
Cardiovascular and metabolic diseases
Spatial

ABSTRACT

The fetal origins hypothesis posits that adverse prenatal exposures, particularly malnutrition, increase the risk of poor adult health. Studies using famine as a natural experiment to test the fetal origins hypothesis present conflicting findings, partly because of data limitations and modeling flaws. Capitalizing on the biomarker data and prefecture-level geographic information from the 2011 China Health and Retirement Longitudinal Study, this study estimates the effects of prenatal exposure to China's 1959–61 famine on later-life risks of cardiovascular and metabolic diseases. Our analysis addresses the problems of measurement error and intrinsic cohort differences that challenge prior studies. We use provincial and prefecture-level geographic variations in famine severity, a proxy for prenatal malnutrition, for model identification. We construct instrumental variables from geocoded newspaper archive data to adjust for measurement error in famine exposure. We find that estimates of the famine effects are highly sensitive to the choices of health indicators, measures of famine severity, and regression model specifications. Overall, we find little evidence supporting the fetal origins hypothesis. In fact, it appears that prenatal exposure to famine reduces later-life disease risks in certain cases. We interpret this finding as evidence of mortality selection among the famine survivors at work. We conclude that using famine as a natural experiment in itself does not guarantee correct statistical inference about the long-term health impacts of prenatal malnutrition when other analytical challenges remain unresolved.

1. Introduction

The controversial fetal origins hypothesis (Barker, 1990, 1995a,b) conjectures that prenatal exposure to an adverse environment, in particular to malnutrition, may "program" the fetus to develop particular metabolic characteristics, likely through environmental effects on the epigenome. Such developmental changes may persist over the life course and increase risks of cardiovascular and metabolic diseases in middle and later ages. The life course perspective of the fetal origins hypothesis, as well as its recent variants such as the developmental origins hypothesis (Bateson et al., 2004; Gluckman et al., 2008, 2005), has made it an attractive theoretical and analytical framework to researchers. However, supportive empirical evidence is uneven, particularly in the area of the long-term health effects of prenatal malnutrition.

Early efforts to test the fetal origins hypothesis using observational data often failed to control for other prenatal confounders (Barker, 1995a,b; Barker and Osmond, 1986). These studies also rely heavily on low birth weight as a crude proxy for fetal malnutrition, even though maternal malnutrition during gestation may induce later-life disease without affecting birth weight (Roseboom et al., 2001). More importantly, the observed association between birth weight and health outcomes in later life could reflect many unobserved joint determinants such as genetic, socioeconomic, and environmental factors. Failing to control for such factors can introduce omitted variable bias and preclude causal inference (Almond and Currie, 2011; Paneth and Susser, 1995; Portrait et al., 2011; Song, 2013a).

To adjust for potential confounders more effectively, researchers have increasingly used exposure to prenatal famine as a natural experiment in studies of long-term health effects of prenatal malnutrition (for a recent review, see Lumey et al., 2011). Well-known famine examples include 19th-century crop failures in Sweden and Finland, the Siege of Leningrad of 1941–44, the Dutch

* Corresponding author.
E-mail address: xuhongw@umich.edu (H. Xu).

http://dx.doi.org/10.1016/j.socscimed.2015.11.028
0277-9536/

Hunger Winter of 1944–45, the Chinese Great Leap Forward famine of 1959–61, and the Bangladesh famine of 1974. Because exposure to famine is beyond the control of most individuals, regardless of their genetic traits, personality, or socioeconomic status, the process governing an individual's prenatal exposure to famine-induced malnutrition is arguably exogenous and resembles random assignment. Therefore, casual effects of prenatal malnutrition on adult health can be inferred by comparing two similar subpopulations that differ in the famine exposure.

However, famine studies have produced conflicting evidence about the effects of prenatal exposure on adult health. For example, in comparisons of cohorts born before or after a famine, prenatal malnutrition has been related to higher older age mortality rates in the 1846–47 Dutch Potato Famine (Lindeboom et al., 2010), but not in the 1866–68 Finnish famine (Kannisto et al., 1997), the Dutch Hunger Winter (Painter et al., 2005), or the 1959–61 Chinese famine (Song, 2009). Even studies of the same famine have produced mixed findings on similar health outcomes. For example, in studies focused on the Siege of Leningrad, Koupil et al. (2007) reported significantly increased risks of cardiovascular diseases (CVD) and mortality in adult life among those exposed to famine in childhood, while Stanner et al. (1997) found no elevated risk for CVD among individuals exposed to famine either in utero or infancy. Similarly, in studies of the Dutch Hunger Winter, Ravelli et al. (1998) found an association between prenatal famine exposure and insulin resistance, but de Rooij et al. (2006) failed to do so. In their review of 30 studies of the relationship between prenatal famine and adult health, Lumey et al. (2011) found a lack of consistent associations for most measures of adult health, except for adult body size, diabetes, and schizophrenia.

These inconclusive findings may reflect incomparable analytic strategies among these studies. For instance, while the most common estimation approach is simple cohort difference (SCD) in health between individuals born during versus after/before a famine, some studies use an approach known as difference-in-differences (DID) to exploit spatial variation in famine exposure in addition to temporal (cohort) variation. The SCD approach hinges on stronger assumptions and presumably yields less robust results than the DID approach (see details below).

In this study, we seek to advance the literature by examining the effects of prenatal and infancy exposure to the 1959–61 Chinese famine, also known as the Great Leap Forward (GLF) famine, on adult cardiovascular and metabolic disease risks. We choose the GLF famine for a case study because of its greater magnitude relative to other famines in terms of duration (three years), geographic scope (pandemic as opposed to endemic), and level of damage (16.5–30 million excess deaths with a mortality rate of over 3.0%) (Song et al., 2009; Susser and St Clair, 2013). Our analysis tackles two empirical challenges of prior studies: (1) measurement errors related to famine exposure and adult health outcomes, and (2) inappropriate estimation strategies. We overcome the first challenge by drawing on biomarker data from the 2011 wave of the China Health and Retirement Longitudinal Study (CHARLS), matching CHARLS respondents to a variety of historical famine data, including fine-grained geographical variation in famine severity. We address the second challenge by assessing results obtained from four discrete estimation strategies – SCD, deviation from cohort trend (DCT), DID, and instrumental variable (IV) – which vary in their ability to address measurement error and to make causal inference. We provide a sketch of the GLF famine in the next section before reviewing the related literature.

2. Background of the GLF famine

From 1958 to 1961, the Communist Party of China (CPC) launched a massive campaign, known as the Great Leap Forward (GLF), mobilizing the entire country to adopt radical economic and social policies to rapidly transform China from a predominantly agrarian society to an industrialized socialist economy through a Soviet-style high investment in heavy industry, supported by agricultural collectivization. To appeal to their zealous superiors in the CPC, and avoid being labeled "anti-revolution," local cadres began to make fictitious high-yield agricultural reports of grain output to the *People's Daily*, the CPC's official newspaper (Kung and Chen, 2011). The first such instance was on June 8, 1958, when the front-page headline of the *People's Daily* reported that a People's Commune in Henan Province achieved a significantly higher than average wheat yield of 2105 catties per *mu* (1 catty = 1/2 kg; 1 *mu* = 1/6 acre). This exaggeration was topped the next day when the *People's Daily* reported that another commune in Hubei Province harvested an average of 2357 catties of wheat per *mu*. Quickly, other regions throughout the country began to over-report grain yields. From June to September, more than 800 false reports (based on our calculation) of abnormally high grain yields were published in the *People's Daily*. At the end of 1958, the national grain production was reported to be 375 million metric tons (MMT), roughly double the yield of 1957. Subsequent verification in 1961, however, placed the actual 1958 yield at 200 MMT (Bernstein, 1984).

These grain yield exaggerations led to food shortages in several ways. Top political leaders, believing China was facing a grain surplus rather than a shortfall (Bernstein, 1984; Yang, 1996), raised compulsory procurement levels — amounts that collectives must deliver to the state (Ashton et al., 1984; Bernstein, 1984). The total grain procurement in the 1958 grain year was 22.3 percent higher than that in 1957 (Yang, 1996). In addition, new policies were implemented to divert labor and resources from agriculture to fruitless projects such as the so-called backyard furnace movement in later 1958 and to reduce sown acreage in 1959 (Ashton et al., 1984). Together, these changes resulted in sharp declines in grain production, and rural villages suffered from severe food shortage after compulsory procurement to support urban and industrial growth. Coupled with other manmade and natural devastating factors, the resulting GLF famine of 1959–61 caused an estimated 16.5 to 30 million excess deaths, depending on the data sources, underlying assumptions, and methods of estimation employed (Ashton et al., 1984; Banister, 1987; Coale, 1984; Peng, 1987; Yao, 1999).

3. Studies of the GLF famine and limitations

In prior studies of the long-term effects of the GLF famine on physical health, the most frequently examined health outcome is adulthood nutritional status derived from anthropometric measures. A key data source in this research is the China Health and Nutrition Survey (CHNS), a longitudinal study of Chinese households in nine provinces since 1989. For example, using data from the 1991 wave of the CHNS, Chen and Zhou (2007) found that those born in 1959 and 1960 attained shorter height in adulthood than those born between 1963 and 1967 (i.e., after the famine). The negative association between prenatal famine exposure and adult height was also confirmed in the 1989 wave (Meng and Qian, 2009) and the pooled 1989–93 waves (Fung and Ha, 2010). However, using the 1989–97 CHNS data, Gørgens et al. (2012) reported that the young female famine cohort (born 1957–61) grew about 2 cm taller in adulthood than did the control cohorts (born 1938–47 or 1962–71) – a finding attributed to a positive selection among famine survivors. Findings on BMI-related measures are even more inconclusive. Studies comparing adults from the famine and post-famine birth cohorts report the famine cohort has no significant difference in BMI (Meng and Qian, 2009), borderline significantly