

September 1, 2019

*Lost and Found* and the Estimate Debate over the “Missing Girls”

For the book *Lost and Found*, we focus more on the notion of mutual noncompliance and the variation in preference for sons than on the exact estimate of hidden girls. This is because there is no a definitive number for the missing girls. Still, the 2010 census shows about 10 million additional girls. Some scholars and demographers observe the same number of additional girls hidden in the 2010 census (10 million), but they have a different interpretation of the same data (Cai 2013, 2017). This sparks a healthy debate. However, we believe the historical, statistical and qualitative evidence we present in *Lost and Found* supports what we observe in the census data and on the ground that there may be more hidden girls in the population than the previous literature suggests. Indeed, we are challenging the basic assumptions behind the numbers. Previous studies tend to assume a relatively consistent implementation of birth policy and historically deep-seated preference for sons. Thus, there is no reason for a large number of villagers to keep their daughters. From this perspective, most of the 10 million additional girls in the 2010 Census must be a statistical error. From our perspective, we believe the evidence suggests many of these girls maybe hidden (i.e. exist). In the book, we suggest close to half of the estimated missing girls may exist and, even if less than half the estimated 20 million missing girls exist, this is still a significant number and supports the mutual noncompliance argument.

The debate over the 2010 census starts with our admittedly crude estimates of the 2010 census in 2016 and the *China Quarterly* responses from Cai Yong (2017) and our response (Shi and Kennedy 2017) in the same issue. In the 2019 book *Lost and Found*, we focus on our 2017 response rather than the full debate. The reason was that we wanted to focus more on the explanation for unregistered births than the numbers debate. However, after publication of the book, we see that number debate deserves greater attention.<sup>1</sup>

While there are no definitive numbers on the missing (hidden or completely gone) girls, it is important to attempt to get the closest estimate possible using the available data. In general, Shi and Kennedy (2106) overestimated the number of hidden girls in our crude estimate of the 2010 census. Based on Cai Yong’s (2017) response to our article and his constructive suggestions we believe the final estimate is closer to 10 million (under 10 million) hidden girls who show up in the 2010 census due to late birth registration. Indeed, Cai Yong (2013) found 10 million additional girls in the 2010 census and we also see the same number in our final crude estimate (2019). However, he believes the 2010 census dramatically overestimates the number of additional or hidden girls, while we believe this may reflect the more wide spread practice of under reporting births. The question is how many are really additional girls (hidden) or just a statistical artifact (still truly missing).

Cai (2013, 2017) uses national primary school education data to test whether or not the girls show up in primary school enrollment figures. Primary school education is basically universal in

---

<sup>1</sup> In the book, we use census data and qualitative interviews. One chapter focuses on historical and current census data and two chapters focus on the interviews. The qualitative data is meant to demonstrate the process and how mutual noncompliance works. Also the interviews show how the preference for sons varies among villagers. Of course, the interviews are not meant to be representative.

China. Therefore, this is a good and a *valid* statistical test. Examining grade cohorts and enrollment sex ratios with corresponding sex ratio at birth from the 1980s, 1990s and 2000s, Cai (2013, 2017) shows that most of these girls are still truly missing. Of course, Cai also admits that there are problems with the education data and even if the numbers are inflated the data shows the girls are still truly missing. In other words, the education data supports the claim that the 2010 census drastically overestimated the hidden girls. However, we (2017, 2019) show that there is sufficient evidence to question the *reliability* of the education statistics including rural students who attend, but are not reported and those who do not attend at all especially at the junior high school level (most girls tend to show up in the census in their early and late teens). In addition, there were issues regarding the inclusion (or exclusion) of migrant children in the education data before 2010. Therefore, we think there are enough problems with the national education data including reporting of enrollments, graduates and migrant children that may not rule out many of additional girls from the 2010 census.

If Cai (2017) is correct, the number of hidden girls in the 2010 census may be a few million (under 5 million) and if Kennedy and Shi (2019) are correct the number of hidden girls maybe around 8 million (in the book we say closer to half of the reported missing girls). Still, the actual number remains unknown. Even if the number of hidden girls is somewhere in the middle, this is still more than the previous literature suggests and supports the mutual noncompliance argument.

Thus, *Lost and Found* is about mutual noncompliance, late birth registration and the variation in preference for sons and the value of daughters. Regarding the census numbers and estimates, we hope that readers will look at the numbers in the historical, cultural and political context (i.e. reading the numbers within the context the whole book). We believe mutual noncompliance is a possible explanation for late registration and many of additional girls that appear in the 2010 census.

John James Kennedy  
Director of the Center for East Asian Studies  
University of Kansas

Shi Yaojiang  
Director of the Center for Experimental Economics in Education  
Shaanxi Normal University

Cai, Yong. "China's new demographic reality: Learning from the 2010 census." *Population and development review* 39, no. 3 (2013): 371-396.

Cai, Yong. "Missing Girls or Hidden Girls? A Comment on Shi and Kennedy's "Delayed Registration and Identifying the 'Missing Girls' in China"." *The China Quarterly* 231 (2017): 797-803.

Shi, Yaojiang, and John James Kennedy. "Missing Girls, Indirect Measures and Critical Assumptions: A Response to Yong Cai's Comments." *The China Quarterly* 231 (2017): 804-810.

Kennedy, John James, and Yaojiang Shi. *Lost and Found: The Missing Girls in Rural China*. (Oxford University Press, 2019).