THE LIMITS AND CONSEQUENCES OF POPULATION POLICY: EVIDENCE FROM CHINA'S *WAN XI SHAO* CAMPAIGN

Kimberly Singer Babiarz, Paul Ma, Grant Miller, and Shige Song

Online Appendix

Appendix 1: Investigation of Female Underreporting in China's 1988 "Two-Per-Thousand" National Survey of Fertility and Contraception

Following a large body of research, we assume that absent sex selection, the sex ratio at birth in modern China should approximately be the naturally occurring rate of 105-106 males per 100 females (Johansson and Nygren 1991). Without the technological ability to identify and selectively abort female fetuses (introduced around the time of the One Child Policy (Chen et al., 213), high sex ratios at birth reflect either under-reporting of children born alive that died early in life or under-enumeration of children living at the time of the survey. The interpretation of these unreported female births is central to our paper. We interpret these "missing girls" to reflect differential rates of infant/child female death, but if the majority of unreported girls were living but simply uncounted, then our interpretation would be incorrect.

We use several methods to investigate the extent to which unreported girls lived beyond infancy as unregistered and unenumerated children. We first carefully consider the potential role of adoption. We then test empirically for systematic under-reporting of living children who could have been adopted-out, or otherwise hidden from enumerators, using three approaches, modifying methods originally developed to evaluate the quality of the 1982 "One-Per-Thousand" national fertility survey by Ansley Coale and Judith Bannister, and directly comparing the 1988 "Two-Per-Thousand" national fertility survey to the 1982 survey (which is generally considered good quality) (Banister 2004; Bhrolchain and Dyson 2007; Coale 1991; Coale and Banister 1994).

Adoption and Survey Design

Before applying established demographic methods for assessing under-reporting of living girls, we first briefly consider how the design of the "two-per-thousand survey" (and enumerator instructions) handles adoption – a specific potential form under-reporting.[1] Survey enumerators were instructed to ensure that adopted children ("adopted-in") were not listed in pregnancy histories as "own children" – and also to ensure that children given up for adoption ("adopted-out") were included in these histories. To accomplish this, the survey included cross-validation measures designed to explicitly handle adoptions in this way (SFPC, 1988) .[2] Although we are of course unable to verify how enumerators conducted fieldwork in practice, systematic under-reporting of children adopted-out (along with other types of under-reporting) would be captured by our analyses below.

[1] If boys adopted into families are reported in our survey's fertility records, or if girls who were given up for adoption are not reported, then the sex ratios that we compute would be inflated.

[2] Specifically, before asking questions about each pregnancy, enumerators were instructed to ask how many "own children" were currently living with the respondent, how many were not living with the respondent, how many had been given up for adoption, and how many had died. Summing across these answers, enumerators were then to calculate the number of pregnancies resulting in live birth – including children "adopted-out." If this cross-validation exercise yielded discrepancies, "interviewer should probe for omissions, twins and multiple births, or to see if adoptive children were listed as own children, etc." (SFPC 1988).

Empirical Assessment of Under-reporting

Method 1

First, following Coale and Bannister (1994), we investigate the extent to which possibly unreported female births in the 1988 "Two-Per-Thousand" survey 're-appear' as adult women in China's population censuses, focusing on those births most likely to be underreported. We compare sex ratios at birth (number of male births for each 100 female births) for each birth cohort reported in the 1988 fertility survey with sex ratios for the same birth cohorts as reflected in the 1% micro samples of the 1982 and 1990 censuses. From cross-sectional census microsamples, we reconstruct sex ratios at birth by adjusting population counts for age- and sex-specific mortality rates, using a reverse survival method.[3] We find that sex ratios at birth in the 1988 fertility survey are consistent with mortality-adjusted sex ratios observed among the same birth cohorts in both the 1982 and 1990 population censuses (Appendix Figures A1-A2 and Appendix Table A1).

To the extent possible, we also investigate the degree to which higher birth order girls (who may have been alive but disproportionately under-reported in fertility histories) are more likely to appear in later population censuses than higher birth order boys. Specifically, we use the same approach described above, stratifying by both birth order sibship sex composition.[4] Due to data requirements, we focus on birth cohorts born between 1975 and 1979 in the 1990 census, adjusting for mortality using birth order-, age-, and sex- specific mortality rates derived from the 1988 "Two-Per-Thousand" survey.[5] We find that girls born at higher parities and with no older brothers – precisely the circumstances under which sex selection is predicted to be strongest – are not more likely to re-appear as adults in future censuses (i.e., we find no evidence of differential under-reporting by birth order and sex composition of previous births) (Appendix Figure A3 and Appendix Table A2).

Method 2

Second, following Coale (1984) (13), we use the 1988 "Two-Per-Thousand" fertility survey to calculate age-specific rate at which women deliver male and female babies in each year. We then apply these fertility rates by maternal age and child sex (simultaneously) to age-specific population counts of women reported in population

[3] Mortality rates are derived from three sets of life tables. These are: 1) life tables presented in Coale (1984), which interpolate between the 1964 and 1982 censuses; 2) life tables published in Bannister (1991), which use China's Cancer Epidemiology Study of deaths between 1973-1975; and 3) life tables based directly on the 1982 population census (Jiang et al., 1984). For all mortality rate adjustments, we necessarily assume that age- and sex-specific morality rates were stable over the period of study.

[4] Because birth order is not directly reported in the population censuses, we reconstruct birth order and sibship sex composition for the subset of individuals still living with their parents in census years. Specifically, we use the total number of boys and girls ever born to a parent together with the sex and age of each child reported in the household roster, restricting our sample to households in which all children born to a mother still co-resided with her at the time of the census (i.e., children for whom we know birth order and sex composition of older siblings with certainty). To maintain cells of adequate size (by birth year, birth order, and sibship sex composition), we focus on birth cohorts from the second half of the 1970s (1975-1980).

[5] Mortality rates from the 1988 "Two-Per-Thousand" survey are shown to be consistent with life-table sources in Appendix Figures A1 - A2 and Appendix Table A1.

census microsamples (interpolated between the 1964 and 1982 censuses), yielding an estimate of the total number of boys and girls born in each calendar year. We then compare the estimated number of male and female births implied by these calculations to the actual number of individuals in each birth cohort in the 1982 and 1990 censuses to estimate the degree of underreporting for boys and girls by birth cohort in the 1988 fertility survey. We find that although females are slightly more likely to be unreported than males, the difference in rates remains relatively constant over time – and in fact *decreases* during the late 1970s (Appendix Figure A4). For underreporting of surviving females to confound our main estimates of missing girls, they would need to increase relative to underreporting of males over time.

Method 3

Third, we investigate the consistency of the 1988 "Two-Per-Thousand" survey with its predecessor, the 1982 "One-Per-Thousand" survey (which others have shown to be good quality (Coale 1984)). To do so, we account for demographic changes between survey years by creating a matched sample of women across surveys. Specifically, for every woman in the "One-Per-Thousand " survey, we identify a woman in the "Two-Per-Thousand" survey with the same characteristics,[6] pooling matched observations from both surveys together. We then regress, separately, (1) the reported number of children (male and female combined), (2) the reported number of male children, and (3) the reported number of female children on a dichotomous indicator variable for which survey the observation was drawn from. The results imply that the number of children (male and female combined) recorded in the 1988 "Two-Per-Thousand" survey is 0.026 fewer than in the 1982 survey [95% CI: 0.014 - 0.037]. Analogous estimates by sex imply 0.0164 fewer male births [95% CI: 0.010 - 0.023] and 0.009 fewer female births [95% CI: 0.003 - 0.016] in the 1988 survey. Overall, these results suggest a small degree of underreporting in the 1988 two-per-thousand relative to the 1982 survey. However, because underreporting is, to a small extent, *more* severe for male births than for female births, the implication is that our estimates of sex ratios at birth during the 1970s may be biased *downwards*.

[6] Individuals in each survey are matched using five individual-level characteristics: birth cohort, urban/rural residence, province of residence, educational attainment, and ethnicity. This means that the pooled data set includes an equal number of women from the "One-Per-Thousand" and "Two-Per-Thousand" surveys, with each observation from the "One-Per-Thousand" survey matched to an observation in the "Two-Per-Thousand" survey sharing exactly the same characteristics along these five dimensions.

Appendix 2: Alternative Explanations for Postnatal Sex Selection

In this section we consider two potential mechanisms by which missing girls could be the indirect result of male-biased stopping rules rather than deliberate sex selection. First, couples that have not yet had a son (or the desired number of sons) may wean a daughter sooner than if the child were a son in an attempt to conceive again more quickly (Jayachandran and Kuziemko 2011). If breastfeeding is protective of infant/child health, this early weaning could result in higher infant mortality rates among infant girls (WHO 2000). Second, when male-biased stopping rules are used, girls will, on average, have more siblings than boys (Clark 2000, Jensen 2003), potentially leading to a quantity-quality tradeoff of sorts (Becker 1991).

<u>Early Weaning of Girls</u>

Because sample selection prevents us from directly estimating excess mortality associated with early weaning ("missing girls" are by definition not present in our data, so we have no information about how long they were breastfed), we instead use back-of-the-envelope calculations to approximate the share of missing girls that may be due to early weaning.

We begin by following Jayachandran and Kuziemko (2011) to derive excess infant mortality rates attributable to early weaning during years *prior to the LLF* policy (years during which we find no evidence of postnatal selection or missing girls). Specifically, we calculate the share of children who are breastfed to 12 months (BF^{12}) and the infant mortality rate (IMR) for children in each parity by sibship sex composition cell (Table A8, Rows 2-3). Then, we use existing estimates of the odds ratio of infant mortality given early weaning (OR^{BF}=1.8) (WHO 2000) to calculate the 'breastfed mortality rate' (IMR^{BF}) as $\frac{IMR}{BF^{12}+OR(1-BF^{12})}$ and the 'non-breastfed mortality rate' (IMR^{nBF}) as $OR^{BF}\times IMR^{BF}$. The implied difference in mortality due to breastfeeding (ΔIMR^{BF}) is then $IMR^{nBF} - IMR^{BF}$ for children in each cell (Table A8, Rows 4-6).

We then convert these excess mortality *rates* into *counts* of missing girls due to early weaning *during LLF policy years*. To do so, we must measure the relative frequency of early weaning among girls relative to boys (or the gender gap in breastfeeding). Because sample selection is again an obstacle in directly measuring the gender gap in breastfeeding during LLF policy years, we report calculations using two alternative approaches.

The first assumes that if parents wean girls early in order to try to conceive again more quickly, the gender gap in breastfeeding should roughly be proportionate to the share of couples using male-biased stopping rules (which we estimate and report in Section 5.2.3). Again focusing on years *prior to the LLF policy*, we measure the gender gap in breastfeeding as the differential probability of being breastfed to 12 months between girls and boys ($BreastfeedingGap^{preLLF} = Prob(BF^{12}_{male}) - Prob\left(BF^{12}_{female}\right)$). Next, we construct ratios between the gender gap in breastfeeding and the prevalence of stopping rule use in each parity by sibship sex composition cell (Table A8, Rows 7-9). We then apply these ratios to the prevalence of stopping rule use *during LLF policy years* to obtain the implied breastfeeding gender gap in each cell under the policy ($\widehat{BreastfeedingGap}^{LLF}$, shown in Table A8, Rows 10-11).

The second approach simply ignores sample selection and uses the reported gender gaps in breastfeeding during the LLF period in each parity and sex composition cell directly (Table A8, Row 12).

Finally, we multiply the implied gender gaps in breastfeeding obtained by each of the two approaches with the implied differential mortality rate due to breastfeeding to obtain implied female infant mortality rates due to early weaning during LLF policy years ($\widehat{BreastfeedingMortality}^{LLF} = \widehat{BreastfeedingGap}^{LLF} \times \Delta IMR^{BF}$, as shown in Table A8, Row 13). We multiply $\widehat{BreastfeedingMortality}^{LLF}$ by the total number of births (from Chinese vital statistics (China Statistical Bureau 2000)) to approximate the number of missing girls that could be explained by the gender gap in breastfeeding. As Table A8, Row 14 shows, the first approach implies a total of 65,944 missing girls across all parity by sex composition cells, while Table A8, Rows 15-16 show that the second approach implies 18,555 missing girls.

Overall, these two back-of-the-envelope calculations imply that the gender gap in breastfeeding could roughly explain between 9% and 33% of all missing girls during the LLF period ($\frac{65{,}944}{200{,}029} = 32.9\%$ and $\frac{18{,}555}{200{,}029} = 9.3\%$).

Sibship Size

We also explore the extent to which missing girls could be an indirect result of male-biased stopping rules, sibship size, and a quantity-quality tradeoff of sorts (Becker 1991). If couples use male-biased stopping rules, girls will, on average, have more siblings than boys (Clark 2000, Jensen 2003). As a result, fewer household resources may be invested in girls, potentially leading to excess female infant mortality. We explore this possibility by re-estimating Equation (10), controlling for the eventual sibship size (the total number of children that parents ultimately have). Table A9 shows these results, suggesting that our results are not sensitive to controlling for ultimate sibship size – and suggesting that this mechanism may not explain a meaningful share of the missing girls that we find associated with the LLF policy.