

**Supporting Information for  
“Population Sex Imbalance in China before the One Child Policy”**

**Appendix 1. Analysis of Possible Underreporting**

**Figure S1: Sex Ratio at Birth Calculated from 1988 Fertility Survey and Implied by the 1990 Population Census**

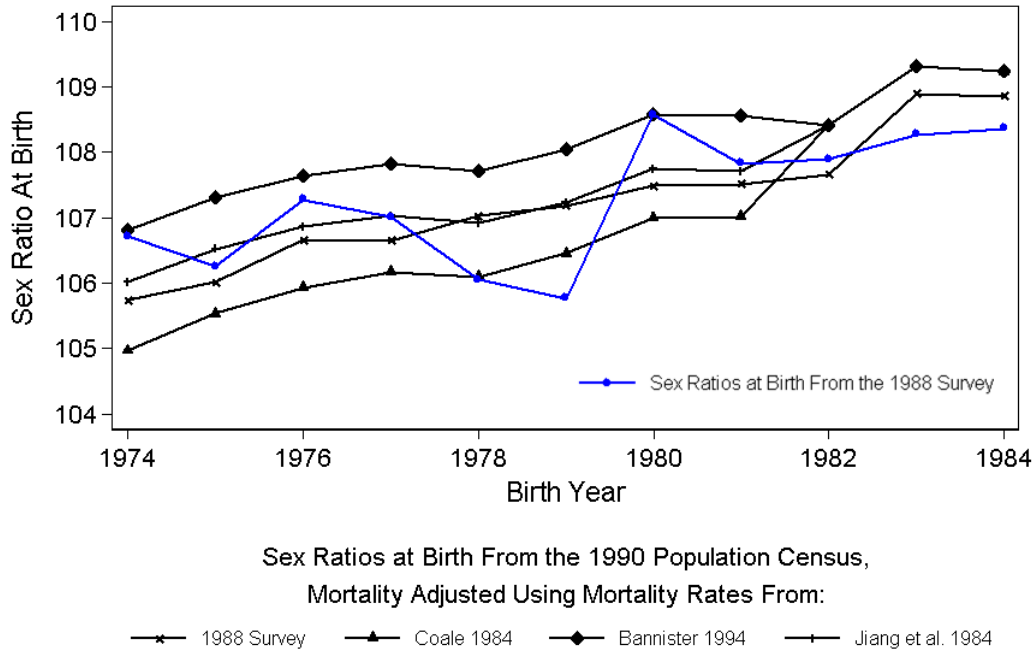


Figure S1 shows sex ratios calculated using the 1988 “two-per-thousand” survey and the 1% sample of the 1990 population census. Population census data adjusted for age- and sex- specific mortality rates using ‘reverse survival’ using mortality rates derived from four sources. These are: 1) mortality rates calculated using the deaths reported in the 1988 “two-per-thousand” survey, 2) life tables presented in Coale (1984), which interpolate between the 1964 and 1982 censuses; 2) life tables published in Bannister (1994), which use China’s Cancer Epidemiology Study of deaths between 1973-1975; and life tables based directly on the 1982 population census (Jiang et al., 1984). For all mortality rate adjustments using life tables, we necessarily assume that age- and sex-specific mortality were stable over the period of study.

**Figure S2: Sex Ratio at Birth Calculated from 1988 Fertility Survey and Implied by the 1982 Population Census**

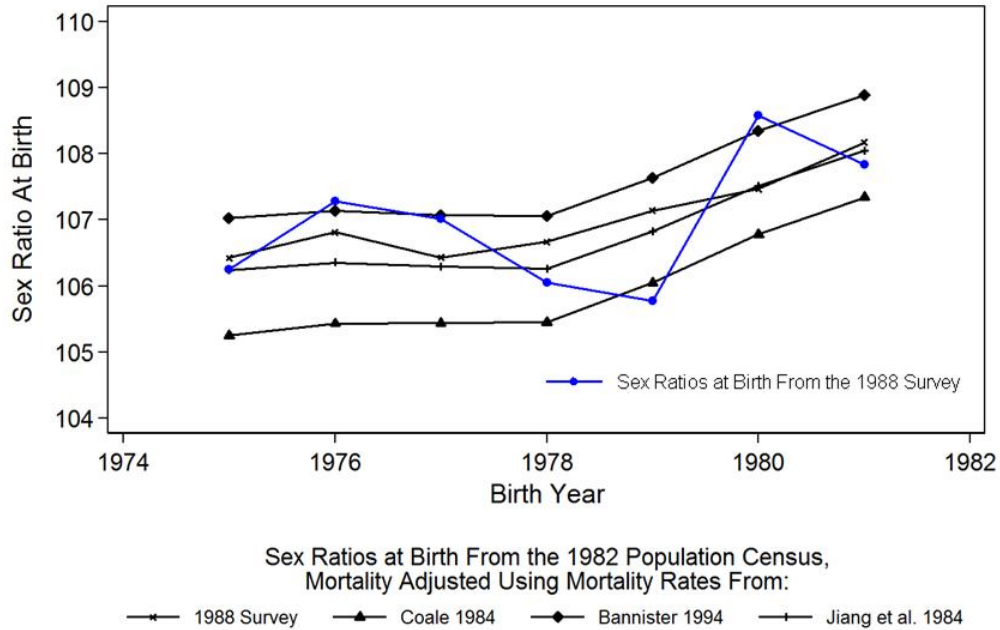


Figure S2 shows sex ratios calculated using the 1988 “two-per-thousand” survey and the 1982 population census. Population census data adjusted for age- and sex- specific mortality rates using ‘reverse survival’ using mortality rates derived from four sources. These are: 1) mortality rates calculated using the deaths reported in the 1988 “two-per-thousand” survey, 2) life tables presented in Coale (1984), which interpolate between the 1964 and 1982 censuses; 3) life tables published in Bannister (1994), which use China’s Cancer Epidemiology Study of deaths between 1973-1975; and life tables based directly on the 1982 population census (Jiang et al., 1984). For all mortality rate adjustments using life tables, we necessarily assume that age- and sex-specific mortality were stable over the period of study.

**Figure S3: Sex Ratio at Birth by Birth Order Calculated from the 1988 Fertility Survey and Implied by 1990 Census Results**

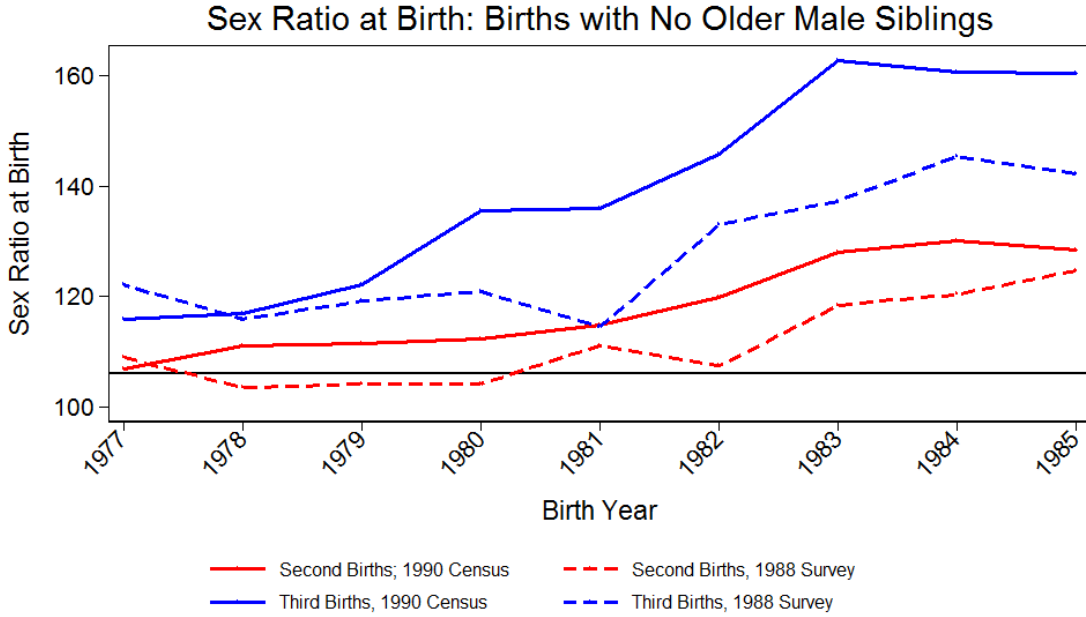


Figure S3 shows parity- specific sex ratios among children born to parents without a surviving male child, calculated using the 1988 “two-per-thousand” survey and the 1% sample of the 1990 population census. Population census data adjusted for age- and sex- specific mortality rates using ‘reverse survival’ using mortality rates derived from the 1988 “two-per-thousand” survey.

**Figure S4: Differential Underreporting of Female vs Male Births by Birth Year**

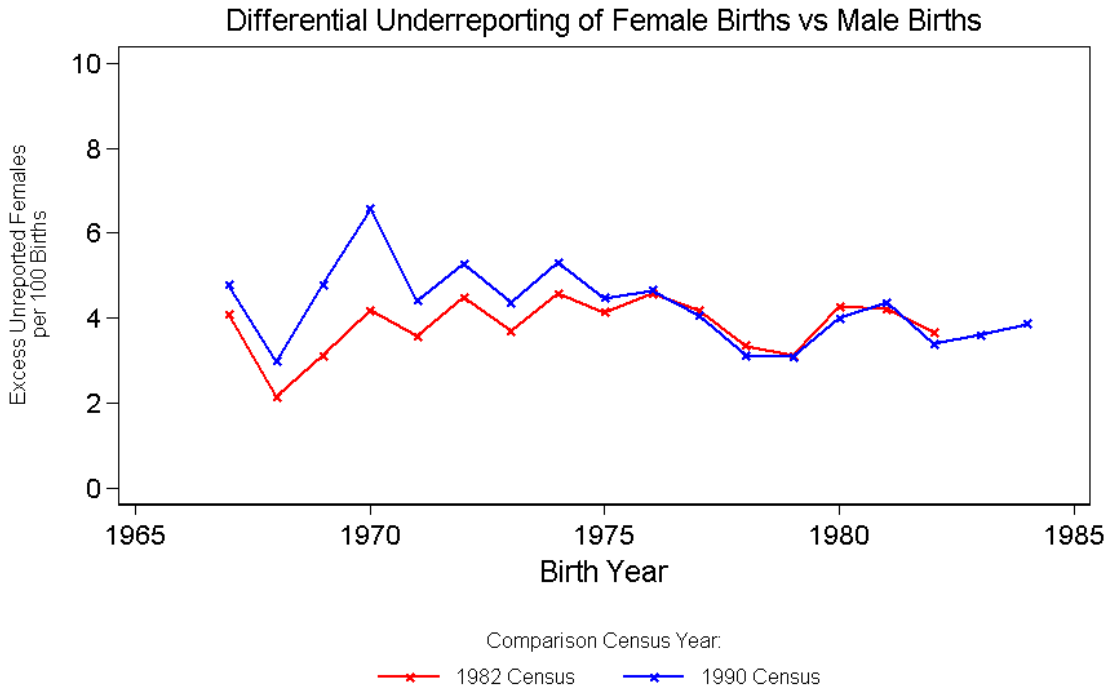


Figure S4 shows the number of male and female births implied by birth rates reflected in the 1988 “two-per-thousand” survey and the number of individuals in each birth cohort enumerated in the 1982 and 1990 population censuses.

PRELIMINARY AND INCOMPLETE - DO NOT CITE

**Table S1: Marginal Probability of a Male Birth, Conditional on Live Birth; Rural and Semi Rural Subsample**

Birth Year	First Birth		Second Birth		Third Birth	
	With Surviving Son	Without Surviving Son	With Surviving Son	Without Surviving Son	With Surviving Son	Without Surviving Son
1965		-0.17%	0.36%	0.96%	0.17%	-0.36%
1966		2.52% ***	2.35% **	0.94%	1.38% *	2.45% ***
1967			..... Reference Year .....			
1968		-0.52%	-0.15%	-1.07%	-0.73%	0.43%
1969		-0.16%	-0.31%	0.59%	0.27%	-0.47%
1970		0.46%	-1.14%	0.01%	0.70%	0.97%
1971		1.01%	-1.92% *	0.35%	0.41%	1.83%
1972		-0.75%	0.17%	0.92%	1.16% *	1.51%
1973		0.25%	-1.65%	1.47%	0.30%	1.12%
1974		-0.57%	1.21%	0.38%	1.30% *	2.30% **
1975		0.27%	-0.10%	0.66%	0.56%	1.51% **
1976		1.57%	1.43%	2.50% *	0.98%	1.93%
1977		0.51%	2.73% *	1.19%	0.72%	4.52% ***
1978		1.73%	-1.27%	0.28%	0.83%	2.88% **
1979		0.42%	1.61%	0.25%	1.72% *	3.20% **
1980		1.75%	4.25% **	-0.46%	1.11%	4.91% ***
1981		1.21%	0.70%	3.23%	2.24%	3.46% *
1982		1.09%	4.22% ***	2.76%	0.21%	6.17%
1983		-2.92%	-4.43% ***	-3.05% *	2.30% ***	5.53% ***

Each cell contains marginal probability that a birth occurring in each year, birth order and sibling sex composition category is male. Marginal probabilities are calculated from coefficients estimated using Ordinary Least Squares linear regressions (estimated separately for births of each order) of an indicator for a male birth on indicators for birth year and the sex composition of older siblings, as well as all two - way interactions between birth year and sibship sex composition. We control for residence in urban area, mother's educational attainment strata, mother's age at marriage, and time-invariant province fixed effects. Huber-White robust standard errors are clustered at the province level. Source: 1988 "two-per-thousand" survey. \* p<0.10, \*\* p<0.05, \*\*\* p<0.001.

PRELIMINARY AND INCOMPLETE - DO NOT CITE

**Table S2: Marginal Probability of a Male Birth, Conditional on Live Birth; Rural Subsample**

Birth Year	First Birth		Second Birth		Third Birth	
	With Surviving Son	Without Surviving Son	With Surviving Son	Without Surviving Son	With Surviving Son	Without Surviving Son
1965		-0.90%	0.66%	0.56%	-0.11%	-0.53%
1966		2.40% **	3.17% ***	2.06%	1.60% *	2.15% **
1967	..... Reference Year .....					
1968		-1.37%	-0.15%	-0.95%	-1.04%	1.01%
1969		-0.33%	0.79%	0.50%	0.34%	-0.72%
1970		0.45%	-1.40%	-0.15%	0.37%	1.83% *
1971		0.36%	-1.62%	0.94%	0.36%	2.09%
1972		-1.17%	-0.74%	1.93%	0.80%	1.43%
1973		-0.55%	-1.78%	1.81%	0.21%	1.07%
1974		-0.97%	1.29%	0.13%	1.32%	3.47% ***
1975		-0.95%	0.02%	0.97%	0.71%	2.44% ***
1976		1.08%	0.77%	3.50% **	1.31%	2.48%
1977		-0.71%	2.15%	1.64%	1.02%	5.74% ***
1978		0.84%	-2.48%	0.86%	0.55%	4.55% ***
1979		-0.33%	1.75%	0.48%	1.71%	5.50% ***
1980		0.96%	4.76% **	1.09%	0.23%	4.52% ***
1981		-0.06%	-0.79%	3.47%	2.73%	4.65% **
1982		0.06%	-1.21%	7.28% ***	4.98% ***	7.24%
1983		-3.78% *	7.25% ***	-2.15%	5.74% ***	1.60%

Each cell contains marginal probability that a birth occurring in each year, birth order and sibling sex composition category is male. Marginal probabilities are calculated from coefficients estimated using Ordinary Least Squares linear regressions (estimated separately for births of each order) of an indicator for a male birth on indicators for birth year and the sex composition of older siblings, as well as all two - way interactions between birth year and sibship sex composition. We control for residence in urban area, mother's educational attainment strata, mother's age at marriage, and time-invariant province fixed effects. Huber-White robust standard errors are clustered at the province level. Source: 1988 "two-per-thousand" survey. \* p<0.10, \*\* p<0.05, \*\*\* p<0.001.

PRELIMINARY AND INCOMPLETE - DO NOT CITE

**Table S3: Marginal Probability of Male Birth Relative Among Third Parity Births With No Previously Born Son**

	Full Sample		Rural + Semi- Rural		Rural Subsample	
	LPM	Probit	LPM	Probit	LPM	Probit
No Son x Parity 3 x Year = 1965	-0.01275 (0.015)	-0.03197 (0.036)	-0.01275 (0.015)	-0.03197 (0.036)	-0.01275 (0.015)	-0.03197 (0.036)
No Son x Parity 3 x Year = 1966	-0.00304 (0.018)	-0.00761 (0.046)	-0.00304 (0.018)	-0.00761 (0.046)	-0.00304 (0.018)	-0.00761 (0.046)
No Son x Parity 3 x Year = 1967	..... Reference Year .....					
No Son x Parity 3 x Year = 1968	0.01194 (0.016)	0.02995 (0.040)	0.01194 (0.016)	0.02995 (0.040)	0.01194 (0.016)	0.02995 (0.040)
No Son x Parity 3 x Year = 1969	-0.01919 (0.018)	-0.04813 (0.045)	-0.01919 (0.018)	-0.04813 (0.045)	-0.01919 (0.018)	-0.04813 (0.045)
No Son x Parity 3 x Year = 1970	0.00604 (0.018)	0.01517 (0.044)	0.00604 (0.018)	0.01517 (0.044)	0.00604 (0.018)	0.01517 (0.044)
No Son x Parity 3 x Year = 1971	0.00881 (0.020)	0.02213 (0.049)	0.00881 (0.020)	0.02213 (0.049)	0.00881 (0.020)	0.02213 (0.049)
No Son x Parity 3 x Year = 1972	-0.00222 (0.017)	-0.00555 (0.042)	-0.00222 (0.017)	-0.00555 (0.042)	-0.00222 (0.017)	-0.00555 (0.042)
No Son x Parity 3 x Year = 1973	0.00008 (0.017)	0.00019 (0.043)	0.00008 (0.017)	0.00019 (0.043)	0.00008 (0.017)	0.00019 (0.043)
No Son x Parity 3 x Year = 1974	0.01293 (0.017)	0.03255 (0.043)	0.01293 (0.017)	0.03255 (0.043)	0.01293 (0.017)	0.03255 (0.043)
No Son x Parity 3 x Year = 1975	0.00875 (0.018)	0.02198 (0.045)	0.00875 (0.018)	0.02198 (0.045)	0.00875 (0.018)	0.02198 (0.045)
No Son x Parity 3 x Year = 1976	0.00324 (0.016)	0.00817 (0.040)	0.00324 (0.016)	0.00817 (0.040)	0.00324 (0.016)	0.00817 (0.040)
No Son x Parity 3 x Year = 1977	0.03862** (0.019)	0.09738** (0.047)	0.03862** (0.019)	0.09738** (0.047)	0.03862** (0.019)	0.09738** (0.047)
No Son x Parity 3 x Year = 1978	0.03151* (0.018)	0.07928* (0.045)	0.03151* (0.018)	0.07928* (0.045)	0.03151* (0.018)	0.07928* (0.045)
No Son x Parity 3 x Year = 1979	0.02935 (0.029)	0.07399 (0.073)	0.02935 (0.029)	0.07399 (0.073)	0.02935 (0.029)	0.07399 (0.073)
No Son x Parity 3 x Year = 1980	0.03440 (0.023)	0.08647 (0.059)	0.03440 (0.023)	0.08647 (0.059)	0.03440 (0.023)	0.08647 (0.059)
No Son x Parity 3 x Year = 1981	0.01061 (0.034)	0.02674 (0.085)	0.01061 (0.034)	0.02674 (0.085)	0.01061 (0.034)	0.02674 (0.085)
No Son x Parity 3 x Year = 1982	0.01408 (0.059)	0.03604 (0.150)	0.01408 (0.059)	0.03604 (0.150)	0.01408 (0.059)	0.03604 (0.150)
No Son x Parity 3 x Year = 1983	-0.04987*** (0.014)	-0.12518*** (0.034)	-0.04987*** (0.014)	-0.12518*** (0.034)	-0.04987*** (0.014)	-0.12518*** (0.034)
No Son x Parity 3 x Year = 1984	0.0867*** (0.025)	0.2217*** (0.063)	0.0888*** (0.024)	0.2266*** (0.060)	0.0647** (0.024)	0.1667*** (0.061)

Each cell contains marginal probability that a third- or higher-order birth occurring in each year, and sibling sex composition category is male, relative to comparable births in the reference year 1969. Marginal probabilities are coefficients estimated using Ordinary Least Squares linear regression or probit regression (as specified in column headers) of an indicator for a third- or higher-order male birth on indicators for birth year and the sex composition of older siblings, as well as all two-way interactions between birth year and sibship sex composition. We control for residence in urban area, mother's educational attainment strata, mother's age at marriage, and time-invariant province fixed effects. Huber-White robust standard errors are clustered at the province level. Source: 1988 "two-per-thousand" survey. \* p<0.10, \*\* p<0.05, \*\*\* p<0.001.

**Table S4: Sex Ratios at Birth for All Births, 1988 "Two-Per-Thousand" Survey, 1990 and 1982**

Year	1988 "Two-Per- Thousand" Survey	1990 Census				
		Not Adjusted for Mortality	1988 Survey	Adjusted for Mortality Using:		
			Coale 1984	Banister 1994	Jiang et al. 1984	
1974	106.72	105.24	105.75	104.97	108.07	106.02
1975	106.25	105.74	106.03	105.54	108.67	106.53
1976	107.28	106.19	106.66	105.94	109.11	106.86
1977	107.02	106.49	106.66	106.17	109.37	107.03
1978	106.05	106.45	107.03	106.10	109.29	106.92
1979	105.77	106.84	107.19	106.45	109.64	107.23
1980	108.58	107.43	107.49	107.00	110.19	107.74
1981	107.83	107.37	107.52	107.02	110.19	107.71
1982	107.89	107.77	107.67	108.42	108.42	108.42
1983	108.28	108.69	108.90	109.31	109.31	109.31
1984	108.37	108.64	108.86	109.24	109.24	109.24
1985	110.87	108.65	108.88	109.22	109.22	109.22

Year	1988 "Two-Per- Thousand" Survey	1982 Census				
		Not Adjusted for Mortality	1988 Survey	Adjusted for Mortality Using:		
			Coale 1984	Banister 1994	Jiang et al. 1984	
1974	106.72	106.12	106.49	99.85	98.67	102.34
1975	106.25	106.18	106.43	105.26	107.03	106.24
1976	107.28	106.31	106.81	105.44	107.14	106.36
1977	107.02	106.27	106.43	105.44	107.08	106.29
1978	106.05	106.19	106.67	105.45	107.06	106.26
1979	105.77	106.71	107.14	106.05	107.64	106.83
1980	108.58	107.35	107.47	106.78	108.35	107.52
1981	107.83	107.83	108.17	107.34	108.89	108.04

Source: 1988 "two-per-thousand" survey. Mortality rates drawn from 1988 "two-per-thousand" survey, and life tables published in by Coale (11), Bannister (14), and Jiang et al. (15).



PRELIMINARY AND INCOMPLETE - DO NOT CITE

**Table S5: Sex Ratios at Birth for Second and Third Order Births with No Older Male Sibling**

Year	Second Births with No Older Male Sibling			Third Births with No Older Male Sibling		
	1988 "Two-Per-Thousand" Survey	1990 Census		1988 "Two-Per-Thousand" Survey	1990 Census	
		Not Adjusted for Mortality	Using Survey-Based Mortality		Not Adjusted for Mortality	Using Survey-Based Mortality
1977	107.54	106.82	105.69	120.87	116.00	114.21
1978	106.46	111.11	112.11	114.01	116.94	117.24
1979	102.31	111.41	111.10	115.59	122.17	120.45
1980	102.96	112.26	113.84	123.67	135.63	134.84
1981	108.62	114.92	114.46	121.67	135.93	131.88
1982	108.66	119.92	117.99	139.26	145.91	142.26
1983	120.19	128.13	125.21	140.53	162.85	154.30
1984	123.66	130.11	129.54	158.41	160.76	158.71
1985	127.88	128.52	128.55	145.29	160.54	156.53

Source: 1990 Population Census of China, 1% sample; 1988 "two-per-thousand" survey.

**Appendix 2: Population-level estimates of missing girls from 1970-1979**

Using the results of our statistical analysis, we calculate the implied number of missing girls as follows. First, we compute the implied sex ratio at birth for each year-parity-sibship sex composition cell for each statistically significant estimate ( $p < 0.05$ ) shown in Table 1:

$$\widehat{SRB}_{yps} = \frac{\widehat{ProbabilityMale}_{yps}}{(1 - \widehat{ProbabilityMale}_{yps})} \times 100$$

where  $\widehat{ProbabilityMale}_{yps}$  is the estimated probability that a birth occurring in year  $y$ , at parity  $p$ , and with sibship sex composition  $p$  is male. Table S7 shows the resulting sex ratios at birth ( $\widehat{SRB}_{yps}$ ). We make the conservative assumption that statistically insignificant estimates imply a biologically-expected sex ratio at birth. Because existing estimates of the biologically-expected sex ratio at birth range between 105 and 106 boys per 100 girls (Grech et al., 2002, Johansson and Nygren, 1991, Sen, 1990), we use 105.5 (results are not sensitive to choices in this range).

Next, we calculate the implied share of girls missing in each year-parity-sibship sex composition cell from the following identity:

$$\frac{SRB^e}{100} = \frac{\widehat{SRB}_{yps}}{100 + n^{missing}}$$

where  $SRB^e$  is the biologically-expected ratio at birth and  $n^{missing}$  is the count of missing girls for each 100 female births. Using this expression, we solve for  $n^{missing}$  and divide by 100 to obtain the share of girls that are missing in each year-parity-sibship sex composition cell:  $\left(\frac{\widehat{SRB}_{yps}}{SRB^e} - 1\right)$ . Table S8 shows these results, assuming  $SRB^e = 105.5$ .

Finally, we weight the cells in Table S8 using the number of girls born in each cell as a share of the overall births observed in a given year (using counts shown in Table S9). This yields year-specific rates of missing girls. We scale these year-specific rates by the total number of births across China in each year (Online, 2016), summing across years 1970-1979 to obtain the total number of missing girls during the 1970s implied by our analysis (Table S10).

To calculate the number of missing girls implied by Figure 1, we apply the same procedure described above using unadjusted sex ratios at birth in each year-parity-sibship sex composition cell in our sample.

PRELIMINARY AND INCOMPLETE - DO NOT CITE

**Table S6: Sex Ratio at Birth (Boys per 100 Girls) Implied by Table 1**

Year	First Order Births				
	Births	Second Order Births		Third and Higher Order Births	
		Male Sibling	No Male Sibling	Male Sibling	No Male Sibling
1970	105.50	105.5	105.50	105.50	105.50
1971	105.50	96.32	105.50	105.50	105.50
1972	105.50	105.5	105.50	110.09	105.50
1973	105.50	105.5	105.50	105.50	105.50
1974	105.50	105.5	105.50	110.47	114.74
1975	105.50	105.5	105.50	105.50	111.65
1976	105.50	105.5	105.50	105.50	105.50
1977	105.50	105.5	105.50	105.50	127.20
1978	105.50	105.5	105.50	105.50	116.58
1979	105.50	105.5	105.50	112.62	120.22

**Table S7: Missing girls as a share of female births in "two-per-thousand" survey**

Year	First Birth	Second Birth		Third Birth	
	Male Sibling	No Male Sibling	Male Sibling	No Male Sibling	
1970	0.00%	0.00%	0.00%	0.00%	0.00%
1971	0.00%	-8.70%	0.00%	0.00%	0.00%
1972	0.00%	0.00%	0.00%	4.35%	0.00%
1973	0.00%	0.00%	0.00%	0.00%	0.00%
1974	0.00%	0.00%	0.00%	4.71%	8.76%
1975	0.00%	0.00%	0.00%	0.00%	5.83%
1976	0.00%	0.00%	0.00%	0.00%	0.00%
1977	0.00%	0.00%	0.00%	0.00%	20.56%
1978	0.00%	0.00%	0.00%	0.00%	10.50%
1979	0.00%	0.00%	0.00%	6.75%	13.96%

**Table S8: Number of births in "two-per-thousand" survey**

Year	Female First Births	Female Second Births		Female Third Births		Total number of births observed (male and female)
	Male Sibling	No Male Sibling	Male Sibling	No Male Sibling		
1970	4629	1857	1884	7186	1654	35343
1971	4443	2072	2086	7338	1612	36154
1972	4167	2005	2171	7059	1645	35174
1973	4187	2183	2176	6687	1635	34620
1974	4617	1937	2122	5942	1676	33735
1975	4260	1935	2058	5306	1580	31262
1976	4590	1925	2044	4726	1524	30813
1977	4616	1844	1963	4076	1308	28590
1978	5138	1943	2028	3956	1449	29951
1979	5801	2081	2294	3814	1430	31795

**Table S9: Population-level estimates of missing girls**

<b>Year</b>	<b>Missing girls as share of total observed births</b>		<b>Total Births Occuring Nation-Wide</b>		<b>Population-level estimate of missing girls</b>
1970	0.000%	×	27,877,013	=	0
1971	-0.499%	×	26,199,395	=	-130,652
1972	0.873%	×	26,083,358	=	227,595
1973	0.000%	×	25,041,528	=	0
1974	1.266%	×	22,669,321	=	286,887
1975	0.294%	×	21,376,746	=	62,948
1976	0.000%	×	18,752,772	=	0
1977	0.941%	×	18,073,552	=	170,041
1978	0.508%	×	17,567,268	=	89,246
1979	1.437%	×	17,381,984	=	249,800

955,864

### References

- COALE, A. 1984. Rapid Population Change in China, 1952-1982. Report No. 27. Washington DC  
Committee on Population and Demography, Commission on Behavioral and Social Sciences and  
Education, National Research Council.
- COALE, A. 1991. Excess Female Mortality and the Balance of the Sexes in the Population: An Estimate of  
the Number of "Missing Females". . *The Population and Development Review*, 3, 517-523.
- JIANG, Z.-H., ZHANG, W.-M. & ZHU, L.-W. 1984. The Preliminary Study to the Life Expectancy at Birth for  
China's Population. *International Seminar on China's 1982 Population Census*. Beijing, China.