



On the estimation of female births missing due to prenatal sex selection

Journal:	<i>Population Studies</i>
Manuscript ID	RPST-2020-2023.R1
Manuscript Type:	Research Note
Keywords:	gender, missing births, estimation, Amartya Sen, prenatal sex selection

SCHOLARONE™
Manuscripts

On the estimation of female births missing due to prenatal sex selection

0. Introduction

Over the last thirty years, there has been a growing acknowledgment of the dramatic effect of gender discrimination on the population sex ratio. This recognition is closely linked to the pioneer endeavor by Amartya Sen to put a figure to these “missing women,” using a simple counterfactual method still employed today. The extent of these sex imbalances has led to a closer examination of their mechanisms and consequences, leading in particular to the emergence of new estimates of “missing female births (MiFB),” “missing children,” “missing births” and “missing brides.”

This research note is prompted by a publication by Kashyap (2019) in this journal. Her paper refers to MiFB and offers 40 original estimates of their numbers in 10 countries and for 4 periods. Kashyap’s study relies, however, on an “alternative definition” of missing female births leading to MiFB estimates about half in magnitude of other estimates as we will illustrate further below. After exchanges with the author about estimation procedures, we realized there was a real need to take stock of the concept of MiFB, as it is widely used by statisticians around the world for assessing and monitoring the demographic consequences of prenatal sex selection.

The research note starts with a brief review of the history and development of the concept of MiFB and the way it has been implemented over the years. The next section presents Sen’s original method and an alternative method used in some papers to compute missing female births. It is followed by a longer section in which we put forward three different arguments in support of the original computation procedure. The note ends with a short summary of our findings.

1. History and use of the concept

Demographers started evidencing unusually high levels of female mortality in some areas about fifty years ago. M.A. El-Badry (1969) introduced the notion of excess female mortality while reporting on “a particular pattern of mortality in Ceylon, India and Pakistan where, contrary to general experience, female mortality is higher than that of the males.” Tabutin (1978) also described cases of excess female mortality in 19th century’s Europe. In the following decade, additional local studies, notably from South Asia, showed the extent of contemporary sex imbalances in infant and child mortality and linked it to systematic gender discrimination (D’Souza and Chen 1980; Miller 1984; Harriss 1989). This work led to the first efforts by Amartya Sen to estimate the number of “missing females” in different countries. A study by Jean Drèze and Amartya Sen (1990: 52) estimated at 105 million the number of missing women in 1986. During the next decade, demographers and economists such as Coale (1991) and Klasen (1994) improved on the initial method by providing better estimates of the number of women expected in the absence of gender discrimination and excess female mortality.

Meanwhile, anthropologists and later demographers had also documented the emergence of skewed levels of the sex ratio at birth (SRB) (Jeffery et al. 1984; Hull 1990). It became necessary to distinguish postnatal from prenatal discrimination, the latter being closely linked to sex-selective abortions. The sex imbalances at birth observed in countries such as China, India or South Korea led in turn to the development of the notion of “missing female births” (or “missing girls”). More recently, the excess

of males among young adults in countries that had experienced sex imbalances at birth in a recent past has caused a mounting marriage squeeze and authors have now pointed to the large number of “missing brides” (Ahlawat 2009; Kaur 2013). In what follows, we will focus on the definition of the “missing female births” caused by prenatal sex selection.

2. The definition of MiFB

Before embarking on the technical discussion, let us detail the various demographic and probabilistic notations used in this note:

OB = Number of observed total births (OFB and OMB for observed female and male births),

EB = Number of expected total births (EFB and EMB for expected female and male births) according to Sen’s original method,

OSRB = Observed sex ratio at birth = OMB/OFB,

ESRB = Expected (biological) sex ratio at birth,

MiFB = Number of missing female births according to the original method.

Note that $OFB = OB/(1+OSRB)$ and $OMB = OB \times OSRB/(1+OSRB)$.

The superscript * refers to variables computed with the alternative method (e.g. EFB* and MiFB*).

The concept of missing births responds to a simple counterfactual question: what would be the expected number of female births in the absence of deliberate prenatal sex selection? The number missing is computed as the difference between the expected and the observed numbers of female births.

$$MiFB = EFB - OFB. \quad (1)$$

2.1. Measuring missing female births using Sen’s approach

The method used here is derived from the original estimate by Sen in his work on missing women and depends on the computation of the expected figure for females. Drèze and Sen (1990) computed for instance the expected female population as the number of women a country would have if its sex ratio was the same as in a region devoid of sex discrimination, given its actual male population (see page 52). Applied to births (rather than to populations), and using a biological SRB (rather than a population sex ratio), the expected female births represent a counterfactual derived from the observed number of male births and the expected biological SRB as shown in the formula below:

$$EFB = OMB/ESRB. \quad (2)$$

It follows that

$$MiFB = OMB/ESRB - OFB. \quad (3)$$

2.2. Measuring missing female births using the alternative approach

This method for computing MiFB has often been applied in the literature without any discussion or presentation. This may explain why other authors have also applied their own methods without describing the procedure followed. Kashyap (2019), however, spells out an alternative method of estimation in which the biological SRB is applied to the total number of births rather than to the number of male births (page 63). This alternative procedure can be summarized thus:

$$EFB^* = OB / (1 + ESRB). \quad (4)$$

It follows that

$$\text{MiFB}^* = \text{EFB}^* - \text{OFB} = \text{OB} / (1 + \text{ESRB}) - \text{OFB} = \text{OB} / (1 + \text{ESRB}) - \text{OB} / (1 + \text{OSRB}). \quad (5)$$

The implications of the selection of one of these two methods are considerable as they lead to estimates varying by up to 100% (see also Equation (7) below). For instance, when computing the range of sex imbalances at birth in 2010-15, Kashyap finds 1.79 million missing female births in China and 1.53 million in India. Using the same data but Sen's original method, Bongaarts and Guilimoto (2015) find instead 3.48 million and 3.00 million missing female births in China and in India respectively. These estimates are 95% larger than Kashyap's and the same disproportion in estimates is found for all 40 estimates in her paper. Other well-cited papers using the same alternative method (Arnold et al. 2002 and Jha et al. 2006) lead to an identical shortfall in the number of estimated missing female births.

3. Three approaches to understand the logic and the implications of the computational methods

The variations between methods used firstly stem from the assumption of the reduction in births found in Sen's method. The extent of these variations calls for a systematic examination of their respective merits. In this section, we confront the validity of the original and alternative methods through three different approaches. We start with a deterministic approach of estimating missing female births. We then introduce a probabilistic modeling of sex-selective abortions to be compared with missing female births. We finally propose a consistency analysis of these estimates in relation to other estimates of missing females.

3.1. A Deterministic Approach

We examine here the logic behind Sen's original method (hereafter "original method") and the alternative estimation method. We show in particular that the alternative method implies the presence of excess male births that cannot be accounted for by the process of birth reduction intrinsic to sex-selective behavior—an unexpected outcome that users of this estimation method do not discuss.

It is assumed that sex imbalances at birth reflect the net effect of prenatal sex selection affecting primarily female births via sex-selective abortions. In the original method, in the absence of sex selection, if there are OMB male births, the expected number of female births, EFB, is given by applying the expected sex ratio at birth, ESRB, to the observed number of male births as shown in the above Equation (2). The number of missing female births, MiFB, is then given by the difference between the expected and observed numbers of female births in Equation (3). The total number of expected births would then be given by the sum of the observed number of male births and expected number of female births, that is,

$$\text{EB} = \text{OMB} + \text{EFB} = \text{OMB} + \text{OMB}/\text{ESRB}. \quad (6)$$

This formulation assumes that the female births have been reduced through prenatal sex selection. In contrast, the male population has not been affected by prenatal selection and serves therefore as a benchmark to derive the expected number of female births.

In the alternative method summarized by Equation (4) above, the number of missing female births, MiFB*, is given in the above Equation (5) used by Kashyap (2019). The ESRB is here applied to the total number of observed births (OB) for computing EFB* and the observed births are redistributed by sex according to the expected SRB to compute EFB*. Contrary to the original method, the total number of expected births remains unaffected by prenatal sex selection.

A further implication of this formula, ignored in the papers using the alternative method, is that the expected number of male births is smaller than that observed, i.e. there was an overcount of male births which now gets corrected. This approach may be suited in case of *misreporting of sex at birth* (e.g. in case female births are reported as male) as this formula keeps the total number of births as fixed and simply redistributes the births by sex. However, this does not account for *births that are missed* due to sex selection and the resulting reduction in observed births.

Using the definition of OSRB, the formulas for missing female births can be rewritten as:

$$\text{MiFB} = [\text{OB} \times \text{OSRB} / (1 + \text{OSRB})] / \text{ESRB} - \text{OB} / (1 + \text{OSRB}) = \text{OB} \times (\text{OSRB} - \text{ESRB}) / [(1 + \text{OSRB}) \times \text{ESRB}], \text{ and}$$

$$\text{MiFB}^* = \text{OB} / (1 + \text{ESRB}) - \text{OB} / (1 + \text{OSRB}) = \{\text{OB} \times (\text{OSRB} - \text{ESRB}) / [(1 + \text{OSRB}) \times \text{ESRB}]\} \times [\text{ESRB} / (1 + \text{ESRB})].$$

It can be finally shown that:

$$\text{MiFB}^* = \text{MiFB} \times \text{ESRB} / (1 + \text{ESRB}) \approx \text{MiFB} \times 0.51, \text{ with ESRB in the } [1.04-1.06] \text{ range.} \quad (7)$$

Thus, MiFB^* is nearly half in magnitude of MiFB . This is seen in the comparison of estimates from Bongaarts and Guilmo (2015) and Kashyap (2019) presented in the previous section 2. This may also be illustrated by a simple case study in which we would observe 110 male births against 100 female births in spite of a biological SRB set at 1. The original method leads to an estimate of 10 MiFB . In contrast, the alternative method results in $\text{MiFB}^*=5$. The gap between these estimates stems from the unexplained 5 excess male births (110 observed against 105 expected) derived from the alternative procedure.

3.2. A Probabilistic View

In this section, we offer an internal demonstration of the consistency of the original method with a probabilistic model of sex-selective behavior showing that the estimated number of abortions is equivalent to the number of missing female births derived from Sen's method.

In a heterogeneous population (Alcalde-Unzu et al. 2020), couples practicing sex selection represent only a fraction θ of all couples who give births. These couples, when sonless after $(n-1)$ previous births, will sex select female births at parity n till they get a boy. This may lead them to several selective abortions before the birth of a son. Based on this assumption, we derive the number of aborted female births at a given birth order. Note that the demonstration adopts a longitudinal approach by following the birth history of a cohort of women and that it is strictly based on the births, not on pregnancies. The model also assimilates the expected proportion of male births to the probability of conceiving a boy.

Additional probabilistic notations are introduced here, with the subscript n referring to parity:

$$\pi = \text{Expected proportion of male births at any parity} = \text{ESRB} / (1 + \text{ESRB}),$$

$$p_n = \text{Observed proportion of male births at parity } n = \text{OSRB} / (1 + \text{OSRB}),$$

$$\theta_n = \text{Proportion of couples who will abort female births until having a male birth, i.e. proportion of sex selectors at parity } n, \text{ and}$$

$$\text{FA}_n = \text{Number of aborted female births at parity } n.$$

Since we assume that the sex selection only happens in the last parity n , MiFB in this section also stands for the missing female births at parity n . Similarly, $\text{FA}_n = \text{FA}$. The expected and observed SRB at parity n can be rewritten as:

$$\text{ESRB} = \pi / (1 - \pi), \quad (8)$$

$$\text{OSRB}_n = p_n / (1 - p_n). \quad (9)$$

Then Equation (3) can be written in terms of proportion of male births:

$$\text{MiFB} = \text{OB}_n \times (p_n - \pi) / \pi, \quad (10)$$

and in terms of sex ratios:

$$\text{MiFB} = \text{OB}_n \times (\text{OSRB}_n - \text{ESRB}) / [(1 + \text{OSRB}_n) \times \text{ESRB}]. \quad (11)$$

The observed proportion of male births at parity n can be decomposed into three parts:

$$p_n = (1 - (1 - \pi)^{n-1}) \times \pi + (1 - \pi)^{n-1} \times (1 - \theta_n) \times \pi + (1 - \pi)^{n-1} \times \theta_n, \quad (12)$$

$(1 - (1 - \pi)^{n-1})$ is the proportion who have at least one son in the first $(n-1)$ births and hence do not practice sex selection and $(1 - \pi)^{n-1} \times (1 - \theta_n)$ is the proportion with no son among the first $(n-1)$ births and who do not practice sex selection. These first two groups do not practice sex selection and hence have an average of π male births. The third part $(1 - \pi)^{n-1} \times \theta_n$ represents couples who are sex selectors and who will therefore all have a son at parity n .

It follows that

$$p_n = [\pi + (1 - \pi)^n \times \theta_n] \quad (13)$$

Substituting Equation (13) in Equation (9), we can express the observed SRB at parity n using θ_n :

$$\text{OSRB}_n = p_n / (1 - p_n) = [\pi + (1 - \pi)^n \times \theta_n] / [(1 - \pi) - \theta_n \times (1 - \pi)^n]. \quad (14)$$

The denominator represents female births diminished by averted female births among sex selectors. It follows that θ_n can be expressed in terms of male birth proportions as well as of sex ratios:

$$\theta_n = (p_n - \pi) / (1 - \pi)^n = (1 + \text{ESRB})^{n-1} \times (\text{OSRB}_n - \text{ESRB}) / (1 + \text{OSRB}_n). \quad (15)$$

We also know that the number of aborted female births follows a geometric distribution with the “probability of success” (i.e. probability of conceiving a son) at π and the mean of distribution at $(1 - \pi)/\pi$. The number of female abortions can now be computed based on θ_n . We have:

$$\text{FA} = \text{OB}_n \times (1 - \pi)^{n-1} \times \theta_n \times (1 - \pi) / \pi \quad (16)$$

where $\text{OB}_n \times (1 - \pi)^{n-1}$ is the number of couples with no son after $(n-1)$ births, θ_n the propensity to sex select, and $(1 - \pi)/\pi$ the mean number of abortions per such couple. Using Equation (15) to substitute θ_n , we have:

$$\text{FA} = \text{OB}_n \times (1 - \pi)^{n-1} \times [(p_n - \pi) / (1 - \pi)^n] \times (1 - \pi) / \pi = \text{OB}_n \times (p_n - \pi) / \pi = \text{MiFB}. \quad \text{QED.}$$

We can also confirm the same result using sex ratio notations in equation (16):

$$\text{FA} = [\text{OB}_n / (1 + \text{ESRB})^{n-1}] \times \theta_n \times 1 / \text{ESRB}. \quad (17)$$

Replacing θ_n using Equation (15), we can achieve Equation (11). QED

This demonstrates that Sen’s original computation perfectly agrees with a more probabilistic model of sex-selective behavior. In contrast, the alternative method yields figures that do not agree with the estimated number of sex-selective abortions. A generalized model (not shown here) in which couples may sex select irrespective of the sex of their previous births similarly yields to the same correspondence with Sen’s computation.

3.3 A Consistency Argument

We finally examine in this section how far the previous MiFB estimates tally with estimates of missing females. Following Sen's method, the expected female population is computed by applying a "normal" sex ratio, itself derived from the biological SRB and from the sex ratio of survival rates drawn from life tables or from non-discriminating populations. Consistency requires that being the product of both pre- and postnatal sex selection, women missing be the sum of women gone missing at birth in the past and of women missing later in life due to excess female mortality. More precisely, the number of missing females at age a in a given cohort should be equal to the expected number of survivors of girls missing at birth in this cohort added to the excess female deaths experienced by this cohort up to age a . This equality has already been confirmed by Bongaarts and Guilmo (2015) in their estimation of the demographic consequences of pre- and postnatal sex selection in the world.

Let us now return to the previous illustration given in section 3, in which we have 110 male births and 100 female births, and a biological SRB of 1. Let us further assume no mortality at all in childhood, which means that 110 boys and 100 girls will reach age 5. The expected sex ratio at age 5 should be 1 (equal to the expected SRB of 1) and the missing girls at age 5 would then number 10 according to Sen's original method. This figure equals the sum of 10 girls missing at birth and later zero in childhood. In contrast, the alternative method would lead to a discrepancy. The sum of girls missing at birth (MiFB* = 5) and missing due to excess mortality (=0) falls short of the 10 girls missing at age 5. This gap is entirely due to the underestimation of girls missing at birth according to the alternative method. Further simulations with parameters closer to real-life SRB and mortality values would lead to identical gaps between MiFB* and the number of missing females.

In short, the alternative estimation method provides estimates of MiFB that conflict with estimates of the number of females missing later in life. To be consistent, the alternative method would require a change in the way missing females are computed (using the total population as a reference rather than the male population) and would generate an unlikely number of excess males in the population. Since sex selection is implemented through selective abortions and excess female mortality, this excess appears once again illogical.

4. Discussion and concluding remarks

The method modeled on Sen's original computation of missing women remains the only one providing accurate and consistent estimates of the number of missing female births. We have first shown how it captures the demography of fertility reduction inherent in gender-biased abortions without generating any artificial excess male births. We have further shown how these results can be replicated with a probabilistic model of sex-selective female abortions. For that, we have used parameters of sex-selective behavior that are flexible enough to reproduce the sex imbalances at various birth orders observed across the world. We finally showed that this method is the only one to coincide with estimate of females missing later in life and ensures therefore consistency with estimates of excess female mortality.

It seems therefore impracticable to employ an alternative method as it would lead to numerous demographic inconsistencies. The alternative method is based on a set of counterfactuals that entirely misses the process of birth reduction underlying sex-selective abortions. This method results in incorrect estimates of MiFB that seriously understate the impact of prenatal sex selection almost by a factor of 2, a variability level unacceptable for any demographic measurement. The source of the confusion may be linked to the initial notion of "missing women" when a more neutral term such as "gender gap" might have avoided some ambiguities.

Such miscalculations have, however, the potential to blur the seriousness of the issue. They run the risk of confusing journalists, observers, activists and policy makers by generating inconsistency across

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

estimation sources and potential incredulity among data commentators. With 45 million MiFB in the world since 1970 up till 2017 (Chao et al. 2019), this issue is obviously not just a matter of demographic correctness.

For Peer Review Only

References

- Ahlawat, N., 2009. Missing brides in rural Haryana: A study of adverse sex ratio, poverty and addiction. *Social change*, 39(1), pp.46-63.
- Alcalde-Unzu J., Echavarri R. and Husillos J. 2020. An evolutionary model of prenatal and postnatal discrimination against females. *Mathematical Social Sciences* 104: pp.1-14.
- Arnold, Fred, Sunita Kishor, and T.K. Roy, 2002. Sex Selective Abortions in India, *Population and Development Review*, 28(4), pp.759-85.
- Bongaarts, J. and Guilimoto, C.Z., 2015. How many more missing women? Excess female mortality and prenatal sex selection, 1970–2050. *Population and Development Review*, 41(2), pp.241-269.
- Chao, F., Gerland, P., Cook, A.R. and Alkema, L., 2019. Systematic assessment of the sex ratio at birth for all countries and estimation of national imbalances and regional reference levels. *Proceedings of the National Academy of Sciences*, 116(19), pp.9303-9311.
- Coale, A.J., 1991. Excess female mortality and the balance of the sexes in the population: an estimate of the number of "missing females". *Population and Development Review*, Vol. 17(3), pp.517-523.
- Drèze, J. and Sen, A., 1990. *Hunger and public action*. Clarendon Press.
- D'Souza, S. and Chen, L.C., 1980. Sex differentials in mortality in rural Bangladesh. *Population and Development Review*, 6(2), pp.257-270.
- El-Badry, M.A., 1969. Higher female than male mortality in some countries of South Asia: a digest. *Journal of the American Statistical Association*, 64(328), pp.1234-1244.
- Harriss, B., 1989. Differential female mortality and health care in South Asia. *Journal of Social Studies*, (44), pp.1-123.
- Hull, T.H., 1990. Recent trends in sex ratios at birth in China. *Population and Development Review*, 16(1), pp.63-83.
- Jeffery, R., Jeffery, P. and Lyon, A., 1984. Female infanticide and amniocentesis. *Social science & medicine*, 19(11), pp.1207-1212.
- Jha, Prabhat, Rajesh Kumar, Priya Vasa, Neeraj Dhingra, Deva Thiruchelvam, Rahim Moineddin, 2006. Low Male-to-Female Sex Ratio of Children Born in India: National Survey of 1.1 Million Households. *Lancet* 367, pp. 211-218.
- Kashyap, R., 2019. Is prenatal sex selection associated with lower female child mortality? *Population Studies*, 73(1), pp.57-78.
- Kaur, R., 2013. Mapping the adverse consequences of sex selection and gender imbalance in India and China. *Economic and Political Weekly*, XLV(5), pp.37-44.
- Klasen, S., 1994. "Missing women" reconsidered. *World Development*, 22(7), pp.1061-1071.
- Miller, B.D., 1984. Daughter neglect, women's work, and marriage: Pakistan and Bangladesh compared. *Medical Anthropology*, 8(2), pp.109-126.
- Sen, A., 1990. More than 100 million women are missing. *The New York Review of Books*, 37(20), pp.61-66.
- Tabutin, D., 1978. La surmortalité féminine en Europe avant 1940. *Population*, 33(1), pp.121-148.