# The Age Distribution of Missing Women in India

SIWAN ANDERSON, DEBRAJ RAY

Relative to developed countries, there are far fewer women than men in India. Estimates suggest that among the stock of women who could potentially be alive today, over 25 million are "missing". Sex selection at birth and the mistreatment of young girls are widely regarded as key explanations. We provide a decomposition of missing women by age across the states. While we do not dispute the existence of severe gender bias at young ages, our computations yield some striking findings. First, the vast majority of missing women in India are of adult age. Second, there is significant variation in the distribution of missing women by age across different states. Missing girls at birth are most pervasive in some north-western states, but excess female mortality at older ages is relatively low. In contrast, some north-eastern states have the highest excess female mortality in adulthood but the lowest number of missing women at birth. The state-wise variation in the distribution of missing women across the age groups makes it very difficult to draw simple conclusions to explain the missing women phenomenon in India.

We are grateful to Monica Das Gupta and an anonymous referee for helpful suggestions on a previous draft of this paper.

Siwan Anderson (*siwan.anderson@ubc.ca*) is an associate professor of economics at the University of British Columbia and CIFAR Associate. Debraj Ray (*debraj.ray@nyu.edu*) is the Silver Professor, Faculty of Arts and Science, and professor of economics, New York University.

## 1 Introduction

"Missing women" – a concept developed by Amartya Sen (1990, 1992) – refers to the observation that in parts of the developing world, notably in India and China, the ratio of women to men is suspiciously low. Sen translated those skewed sex ratios into absolute numbers by calculating the number of extra women who would have been alive (say in China or India) if these countries had the same ratio of women to men as in areas of the world in which they purportedly receive similar care. Sen estimated that more than 100 million women were "missing", presumably from inequality and neglect leading to excess female mortality.[1]

Research aiming to explain these estimates of missing women in India has primarily focused on excess female mortality at younger ages. In particular, much subsequent attention has been placed on a skewed sex ratio at birth (SRB), which indicates the possibility of sex-selective abortion. For instance, Jha et al (2006) estimate that as many as 10 million female foetuses have been aborted over a 20-year period.[2] A second area of focus is early childhood and the possibility that young girls are systematically less cared for.[3] New estimates from the 2011 Census reveal that the child sex ratio of boys to girls, aged 0 to 6, is at its highest level since Independence. Additional research has demonstrated large geographic discrepancies within India, where these "juvenile" sex ratios are particularly biased in favour of boys in northern states compared to southern ones.[4]

It is not that the excess mortality of females in adulthood is entirely ignored,[5] but the focus is on the prenatal and infant/early childhood stages. Das Gupta (2005) quite fairly summarises the literature when she states that "the evidence indicates that parental preferences overwhelmingly shape the female deficit in south and east Asia".

This paper uses a methodology, developed in Anderson and Ray (2010), to explicitly examine how the missing women in India are distributed across different age groups. Our numbers are estimates of *flows*; i e, excess female mortality per year, and are therefore lower than the stock numbers estimated by Sen and others from overall population ratios. This is of little import as the former can be converted, under some demographic assumptions, into the latter. We essentially apply a variant of the Sen counterfactual to every age group. Briefly, we suppose (for each age group after birth) that the relative death rates of females to males are "free of bias" in developed countries. So for each age group, we posit an "unbiased" death rate for females, one that would obtain if the death rate of males in that state were to be rescaled by the *relative* death rates for males and females (in the same age group) in developed countries. We subtract this unbiased rate from the actual death rate for females, and then multiply by the population of females in that category. We describe the resulting distribution, as well as the aggregate numbers. This exercise enables us to address which age groups and states house the missing women that were identified overall by Sen.

We estimate that a total of more than two million women in India are missing in a given year. Our age decomposition of this total yields some striking findings. First, the majority of missing women, in India die in adulthood. Our estimates demonstrate that roughly 12% of missing women are found at birth, 25% die in childhood, 18% at the reproductive ages, and 45% die at older ages.

One possible reaction to this observation is that excess female mortality is higher in adulthood simply because *overall* mortality, excess or not, is higher in adulthood. Such a reaction would only make sense for older ages, of course. But as we argue below (see especially the discussion in Section 5), this is not the entire story. Relative excess female mortality, by which we mean excess female mortality *per female death*, is highest in the reproductive years and at older ages.

Second, there is significant variation in the distribution of missing women by age group across the states. Only in Punjab are the majority of missing women to be found at birth. There are just two states in which the majority of missing women are either never born or die in childhood (i e, before age 15), and these are Haryana and Rajasthan. Moreover, the missing women in these three states add up to well under 15% of the total missing women in India.

For all other states, the majority of missing women die in adulthood. Among this set of states, there is variation as well. The southern states (Kerala, Tamil Nadu and Andhra Pradesh) have the lowest number of missing women and make up only 10% of the national total. The remaining bulk of the missing women are distributed throughout the rest of the country. One group of states (Madhya Pradesh, Maharashtra, Bihar and Assam), which comprise roughly 37% of the total missing women in the country, have the highest excess female mortality in the country at all ages and the relative differences are largest at adult ages. Another group of states (West Bengal, Orissa, Himachal Pradesh and Uttar Pradesh), which make up roughly 33% of the national total, have lower excess female mortality at the younger ages, but high excess female mortality at adult ages.

The significant state-wise variation in the distribution of missing women across the age groups makes it very difficult to provide a one-dimensional explanation for missing women in India. It is beyond the scope of this paper to specifically identify the different pathways which explain the state-wise variation in the distribution of missing women across India. What the paper makes clear, though, is that excess female mortality in adulthood is as serious a problem in India as missing girls who are never born or die prematurely in childhood.

## 2 Methodology

The estimation of missing women across different age groups and Indian states employs the methodology developed in our previous work; see Anderson and Ray (2010). Let $a$ denote an age group in $\{1,\ldots,n\}$. Use the extra value $a = 0$ to indicate birth. For any age $a \geq 1$, deaths within that group $a$ refers to all deaths between the ages of $a - 1$ and $a$. Let $d^m(a)$ and $d^w(a)$ represent the rate of death of men and women respectively at age $a$. The label ^ denotes these variables in the reference region.[6]

The unbiased death rate for women of age $a$ in the region of interest (here, an Indian state) is defined by:

$$u^w(a) = \frac{d^m(a)}{\hat{d}^m(a)/\hat{d}^w(a)} \qquad \ldots(1)$$

The implicit idea behind the equation is that the relative *male* death rates across the region of interest and the reference region accurately reflects the relative cost of healthcare. Therefore, the unbiased death rate for women "should" bear the same ratio, relative to the prevailing female death rate in the reference region. That is, $u^w(a)/\hat{d}^w(a)$ should be the same as $d^m(a)/\hat{d}^m(a)$, which is exactly what (1) states.

Excess female mortality or "missing women" in the region at age $a$ is then equal to the difference between the actual and unbiased death rates for women, weighted by the number of women in that age group:

$$mw(a) = [d^w(a) - u^w(a)]\pi(a), \qquad \ldots(2)$$

where $\pi(a)$ is the starting population of women of age $a$.

To compute missing women at birth, we compare the SRB in a given state with a presumably "unbiased" SRB, drawn from the reference region. We use a formula analogous to (2) to carry out this computation:

$$mw(0) = [\frac{\sigma(0)}{\hat{\sigma}(0)} - 1]\ \pi^w(0), \qquad \ldots(3)$$

where $\sigma(0)$ is the SRB in the state, $\hat{\sigma}(0)$ the comparison ratio from developed countries, and $\pi^w(0)$ is the total number of female births in the state. In Section 3, we remark on the difficulties of obtaining an appropriate reference SRB.

To obtain the total number of missing women in a given year and state, $mw$, we aggregate across the age groups:

$$mw = \Sigma_{a=0}^{n}\ mw(a) \qquad \ldots(4)$$

So much for absolute numbers. Sometimes we might be interested in the "intensity" of excess female mortality at different ages. Intuitively, this refers to excess female mortality at a particular age relative to total female deaths at that age. This is obviously related to the ratio of the death rate $d^w(a)$ to the unbiased rate $u^w(a)$, or equivalently, we can define relative excess female mortality to be given by

$$\frac{d^w(a)/d^m(a)}{\hat{d}^w(a)/\hat{d}^m(a)}$$

Whenever there are missing women at age $a$, this ratio must exceed *1*. But the extent to which it exceeds *1* is one measure of the "relative importance" of excess female mortality at that age. To be sure, it is only one possible indicator, another being $mw(a)/mw$, the ratio of missing women at age $a$ to the total across all ages.

Throughout the paper, we consider the 16 largest states of India.[7] See Table 1 for a list in order of population size in 2003.

**Table 1: Female Populations for 16 States in India** (2003, in millions)

| State | Female Population | % of India Total |
|---|---|---|
| Uttar Pradesh | 87.7 | 17.0 |
| Maharashtra | 43.7 | 8.5 |
| Bihar | 41.7 | 8.1 |
| West Bengal | 40.9 | 7.9 |
| Andhra Pradesh | 36.9 | 7.2 |
| Madhya Pradesh | 31.8 | 6.2 |
| Tamil Nadu | 30.5 | 5.9 |
| Rajasthan | 26.6 | 5.2 |
| Karnataka | 26.5 | 5.1 |
| Gujarat | 24.0 | 4.7 |
| Orissa | 20.1 | 3.9 |
| Kerala | 17.3 | 3.3 |
| Assam | 11.7 | 2.3 |
| Punjab | 10.9 | 2.1 |
| Haryana | 10.3 | 2.0 |
| Himachal Pradesh | 3.2 | 0.6 |
| India | 515.9 | 100 |

Source: Sample Registration System (Government of India).

## 3 Sex Ratios at Birth

Defining an "unbiased" reference SRB, $\hat{\sigma}(0)$, is a difficult proposition. Ideally, we want as a comparison point the SRB generated by "the same group in the same circumstances", cleansed of any differential treatment for boys and girls. Such a reference point is not available. For example, in the United States (US), the average SRBs for Asian populations is around 1.07. But there is emerging evidence that Asian populations residing in the US may well be practising gender selection at the prenatal stage (Abrevaya 2009; Almond and Edlund 2008).[8] That said, it is unclear at this stage whether the phenomenon is pervasive enough to alter the overall estimates. Certainly, a SRB of 1.07 is by no means an outlier. It is within the average range in developed countries, and it is typical of southern European populations. Nevertheless, despite seemingly systematic racial differences, we should certainly be wary of using the SRBs of Asian populations residing in developed countries as a reference point.

Coale (1991) used a reference SRB of 1.059, which is the median SRB across developed countries. Using this as a reference group for India is potentially problematic: there may be inherent variation in the SRB across race and ethnicity.[9] But we do no damage to the main point of this paper by choosing the Coale reference. If anything, this procedure will overestimate the number of missing girls at birth in India, and our goal is to argue that the number is nevertheless relatively small.

**Table 2: Sex Ratios by Birth by State in India** (various years)

| State | SRB 2000-06 | SRB 1996-2006 | SRB 1997-2002 |
|---|---|---|---|
| Punjab | 1.351 | 1.348 | 1.345 |
| Haryana | 1.283 | 1.223 | 1.202 |
| Rajasthan | 1.123 | 1.089 | 1.055 |
| Orissa | 1.115 | 1.124 | 1.222 |
| Maharashtra | 1.112 | 1.059 | 1.039 |
| Himachal Pradesh | 1.104 | 1.136 | 1.113 |
| Gujarat | 1.103 | 1.098 | 1.133 |
| Andhra Pradesh | 1.095 | 1.060 | 1.098 |
| Kerala | 1.092 | 1.052 | 1.058 |
| Uttar Pradesh | 1.090 | 1.077 | 1.066 |
| Tamil Nadu | 1.076 | 1.054 | 1.000 |
| Bihar | 1.075 | 1.083 | 1.071 |
| West Bengal | 1.062 | 1.050 | 1.050 |
| Madhya Pradesh | 1.059 | 1.055 | 1.028 |
| Karnataka | 1.055 | 1.096 | 1.095 |
| Assam | 1.005 | 1.024 | 1.040 |
| India | 1.081 | 1.074 | 1.075 |

Source: National Family Health Survey (2005-06).

To estimate the SRB for India and Indian states, we use data from the National Family Health Surveys of India (NFHS).[10] In Table 2, we report estimates of SRB for the 16 largest states of India. We have listed states in descending order of the average SRB for 2000-06.

It should be noted right away that the NFHS does not provide us with a very large sample. For the 16 states we consider, approximately 65,000 births are recorded by the survey over 2000-06. Except for 2006, when there was a large drop in the count, the annual count is close to 11,000 births per year, but at the same time the birth numbers are distributed over 16 states. State-level estimates are consequently based on even smaller numbers: between 300 and 1,500 per year, leading to 2000-06 totals that lie between 1,300 (for the smaller states) and 9,000 (for the larger states).

This means that we have to contend with significant variation in the sample SRB. For India as a whole, the ratio runs from a minimum of 1.04 in 1996 to a maximum of 1.11 in 2003. To be sure, some of this might represent a genuine trend, but not all: the 2005 estimate is 1.06, while the 2006 number is far lower and certainly noisy. Owing to this variation we use the average between the years 2000 and 2006 to obtain an estimate of the SRB for 2003, which is 1.08. If anything, this tilts the case against us. Table 2 shows us that other averages that include earlier years, such as 1996-2006, generally lead to lower estimates of the SRB. This is not surprising as a gradual upward creep in the SRB is not inconsistent with the worsening aggregate sex ratios that we see in the Indian census. By using the 2000-06 average, we smooth out the variation and most likely provide an upper bound on the number of missing girls at birth.

The problem of noisy variation is even higher at the state level. But all states fall into one of the following patterns. There are some with a very high SRB, and no amount of noise in the data can hide this. In particular, for Punjab the sample SRB exhibits a minimum of 1.22 in 1999, and a high of 1.55 in 2002. More generally, in states with higher-than-average SRBs (greater than 1.06) – e g, Haryana, Orissa, Himachal Pradesh, and Rajasthan – the SRB is greater than 1.10 for the majority of observations. For Gujarat and Maharashtra, the 2000-06 average exceeds 1.10, though the SRB is less than this value in the majority of observations for these states. Therefore, for these two states our computation of missing women at birth is most likely an overestimate. For Karnataka, the estimated (2000-06) SRB is low, at 1.05, but using additional years of the data we get a higher estimate of 1.09. For this state, we will use the higher estimate of 1.09 (using the data from 1996-2006) so as to avoid an underestimate of missing girls at birth.

At the other extreme, for a few states, we have SRBs that are lower than the national average (smaller than 1.06). It turns out that for these states, the SRB is indeed lower than 1.06 for the majority of year observations. For example, the lowest estimate is for the state of Assam, where, for eight of the 10 years of data, the SRB is less than 1.05.

As discussed earlier, for any state with a SRB below the Coale benchmark ratio of 1.059, we impute no missing women at birth. This is the benchmark we used in our earlier work and the common one used in the literature. We reiterate that it is conceivable that there are other reasons (aside from gender discrimination) which explain why the SRB in the states of India differ from that of developed countries.[11] However, as discussed in Section 2, the choice of a more appropriate benchmark is far from obvious.

## 4 Estimates of Missing Women

We now compute the number of missing women by age for the year 2003, across the different states of India, using the methodology outlined in Section 2. For the reference group, we rely on the estimates from our earlier work using country-level data from the UN and WHO (see Anderson and Ray 2010 for more details). For Indian mortality and population estimates, we use data provided by the Sample Registration System (SRS) of India. As for SRBs, there is large annual variation in the reported death rates using the SRS data. We follow the methodology of the WHO for putting together estimates of death rates by age and gender for the year 2003. We use age-specific mortality rates averaged across 2001-06 and adopt standard demographic techniques to account for the under-reporting of deaths in the data (refer to the Appendix for more details). Age-specific population numbers are obviously more reliable and we use those for the year 2003.

In principle, under-reporting of death rates could affect both the total estimates of missing women, as well as their composition across different age groups. There is, in fact, little doubt that the total numbers are affected, for the simple reason that the under-reporting of deaths for women is higher than that for men. It follows that our correction must impute a larger number of missing women relative to any procedure that takes the reported numbers literally. As for composition, there is little evidence that the patterns of relative under-reporting across gender varies substantially across different age groups. An obvious candidate for excessive under-reporting would be deaths from female infanticide or neglect, but such under-reporting would then effectively manifest itself in greater skewness in the SRB, something that is already in the data. In short, after our correction, the relative distribution of missing women across age groups does not alter significantly.

Column 1 of Table 3 lists the total number of missing women by state in the year 2003. The second column lists the proportion of the total female population (in %) in a given state which is missing. We have listed states in descending order of female proportion missing.

Our estimates suggest that there are more than two million missing women in India in the year 2003.[12] This is a higher number than the estimate in Anderson and Ray (2010) for India, which was 1.7 million for the year 2000. Our larger estimates for the year 2003 are primarily driven by the increase in female population numbers across the years, and the fact that the SRB we use for the year 2003 is higher than the one we used for the year 2000.

**Table 3: Missing Women by Indian State** (2003, in 000s)

| State | Missing | % of Female Population | Proportion of Total |
|---|---|---|---|
| Haryana | 126 | 1.23 | 4.5 |
| Madhya Pradesh | 297 | 0.93 | 10.7 |
| Punjab | 94 | 0.86 | 3.4 |
| Maharashtra | 344 | 0.79 | 12.4 |
| Bihar | 320 | 0.77 | 11.5 |
| Assam | 79 | 0.67 | 2.8 |
| Uttar Pradesh | 570 | 0.65 | 20.5 |
| Himachal Pradesh | 19 | 0.58 | 0.7 |
| Orissa | 115 | 0.57 | 4.1 |
| West Bengal | 216 | 0.53 | 7.8 |
| Karnataka | 119 | 0.45 | 4.3 |
| Rajasthan | 111 | 0.42 | 4.0 |
| Gujarat | 95 | 0.40 | 3.4 |
| Andhra Pradesh | 146 | 0.39 | 5.3 |
| Tamil Nadu | 88 | 0.29 | 3.2 |
| Kerala | 42 | 0.24 | 1.5 |
| India | 2,233 | 0.43 | – |

Sources: National Family Health Survey (2005-06), United Nations, World Health Organisation, Sample Registration System (Government of India).

**Table 4: Missing Women by Indian State and Age Group** (2003, in 000s)

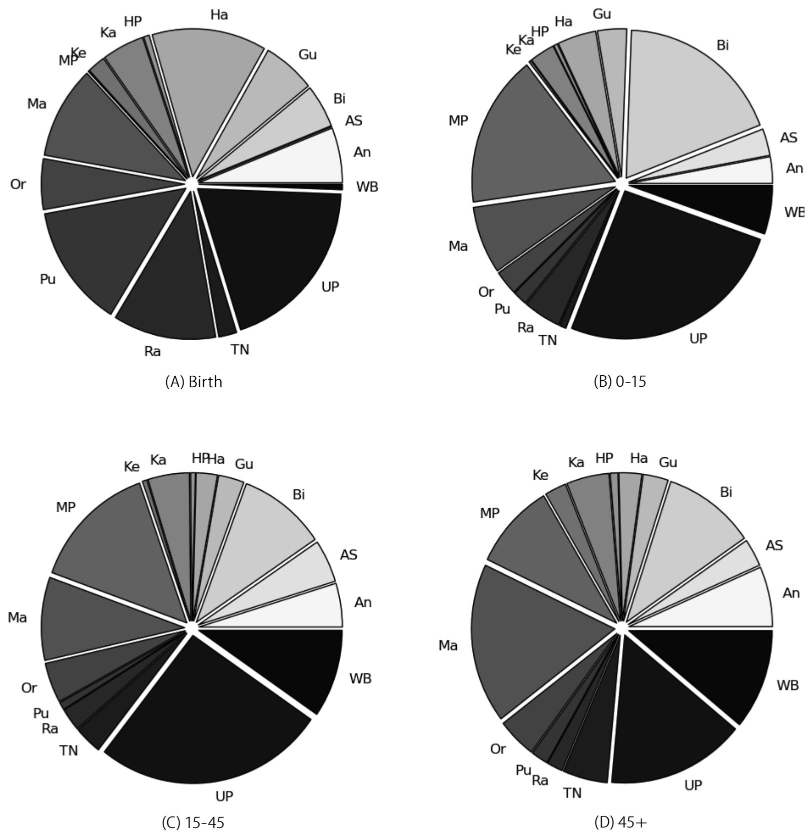
| State | At Birth | | 0-15 | | 15-45 | | 45+ | |
|---|---|---|---|---|---|---|---|---|
| | Number | % | Number | % | Number | % | Number | % |
| Haryana | 53 | 42.4 | 30 | 24.3 | 11 | 9.1 | 30 | 24.1 |
| Madhya Pradesh | 0 | 0 | 117 | 39.4 | 68 | 23.0 | 111 | 37.5 |
| Punjab | 56 | 59.5 | 11 | 12.1 | 4 | 4.3 | 22 | 24.0 |
| Maharashtra | 42 | 12.3 | 52 | 15.1 | 44 | 12.7 | 206 | 59.9 |
| Bihar | 20 | 6.1 | 131 | 40.8 | 47 | 14.8 | 122 | 38.2 |
| Assam | 0 | 0 | 21 | 26.1 | 23 | 28.6 | 36 | 45.3 |
| Uttar Pradesh | 81 | 14.2 | 181 | 31.7 | 125 | 22.0 | 183 | 32.0 |
| Himachal Pradesh | 3 | 14.2 | 3 | 16.6 | 3 | 14.0 | 10 | 55.2 |
| Orissa | 23 | 19.9 | 19 | 16.5 | 21 | 18.4 | 52 | 45.3 |
| West Bengal | 2 | 1.1 | 37 | 17.3 | 46 | 21.5 | 129 | 60.0 |
| Karnataka | 20 | 16.6 | 20 | 16.7 | 22 | 19.0 | 57 | 47.7 |
| Rajasthan | 48 | 43.4 | 30 | 26.8 | 12 | 10.8 | 21 | 19.0 |
| Gujarat | 25 | 25.8 | 23 | 24.2 | 14 | 14.7 | 34 | 35.3 |
| Andhra Pradesh | 25 | 17.1 | 20 | 14.0 | 23 | 15.8 | 77 | 53.1 |
| Tamil Nadu | 8 | 9.5 | 5 | 6.1 | 15 | 17.1 | 59 | 67.2 |
| Kerala | 8 | 20.3 | 2 | 4.4 | 2 | 6.1 | 29 | 69.3 |
| Total | 265 | 11.9 | 558 | 25.0 | 398 | 17.8 | 1,013 | 45.3 |

Sources: National Family Health Survey (2005-06), United Nations, World Health Organisation, Sample Registration System (Government of India).

These missing women are distributed across the country. Expressed as a proportion of state-level female population, excess female mortality is highest in Haryana, Madhya Pradesh, and then Punjab. The states where this proportion is around or below the country average are all the southern states (Karnataka, Andhra Pradesh, Tamil Nadu and Kerala), along with Rajasthan and Gujarat. Kerala displays the lowest proportion of excess female mortality relative to female population.

Of course, state contributions to overall excess female mortality represent a different story. The last column of Table 3 records state-level missing women as a percentage of the total over all 16 states. Now state populations matter, and it is not surprising to see that Uttar Pradesh, Maharashtra, Madhya Pradesh and Bihar are the principal contributors to the 16-state total of missing women. But of these four, the last three are not just in for population alone; they are also "over-contributing" relative to their populations.

**Figure 1: State Distribution of Missing Women, Different Age Categories**




A comparison of the rankings in Tables 2 and 3 is instructive. The ordering of states by sex ratio at birth has little correlation with the ordering by missing women as a fraction of state female population (the Spearman rank correlation coefficient is a shade under 0.15). There is more to missing women than discrepancies in the sex ratio at birth. The discussion that follows brings this point out more vividly.

Table 4 (p 90) computes excess female mortality in different age groups for each state. The first column in each age category is the total number of missing women. The second column under each age category reflects the proportion of total missing women in a given state found in that particular age group. We report states in the same ranking in which they appear in Table 3, in descending order of missing women relative to their female populations.

It is instructive to look, first, at Haryana, Rajasthan and Punjab. They typify what is perhaps the generic viewpoint regarding excess female mortality: that the bulk of missing females are missing at birth or at young ages. Over two-thirds of the missing women are missing by the age of 15. For Punjab, close to 60% of the excess female mortality is at birth, and for Haryana and Rajasthan, the numbers are well over 40%. These are disturbing numbers that reflect the conventional wisdom on missing women. The numbers at birth are not very different from what we found for China (Anderson and Ray 2010: 1275).

But Haryana, Punjab and Rajasthan account for well under 15% of the 16-state total for missing women. Their profile is emphatically not the case for India as a whole. The majority of missing women in India die in adulthood (older than age 15). Roughly 12% of missing women are found at birth, 25% die in childhood, 18% at the reproductive ages, and 45% die at older ages. With the possible exception of Gujarat,[13] the majority of missing women die in adulthood in all the other states.

To be sure, among this latter set of states, there is variation as well. In some states, a large proportion of excess female mortality does occur in childhood, particularly in Bihar, Madhya Pradesh and Uttar Pradesh. More than 20% of the missing women in Assam, Madhya Pradesh, Uttar Pradesh and West Bengal die in the reproductive ages. Finally, more than half of the missing women are at older ages (greater than 45) in Andhra Pradesh, Himachal Pradesh, Kerala, Maharashtra, Tamil Nadu and West Bengal.

Figure 1 describes the distribution of total missing women in India across states. Each of the four panels studies a different age category. To be sure, the larger states such as Uttar Pradesh invariably figure in a prominent way, but relative to their populations (consult Table 1), different states come into their own in different age categories.

Missing women at birth are mainly found in Uttar Pradesh, Punjab, Haryana, Rajasthan and Maharashtra. Punjab and Haryana are exceptionally over-represented relative to their populations. Most of the missing girls who die in childhood are to be found in Bihar, Madhya Pradesh and Uttar Pradesh. Missing women at the reproductive ages in India are mainly found in Bihar, Maharashtra, Madhya Pradesh, Uttar Pradesh and West Bengal. Relative to population shares, this number is highest in Madhya Pradesh and Uttar Pradesh. Finally, missing women at older ages are distributed more uniformly across the states; larger numbers are to be found in Bihar, Maharashtra, Madhya Pradesh, Uttar Pradesh and West Bengal. Relative to population shares, there is over-representation particularly in Madhya Pradesh and Maharashtra.

## 5 State Variation in Excess Female Mortality Rates

One reaction to the preponderance of missing women in adulthood is that overall mortality is highest in adulthood, so it is almost true by definition that a large fraction of missing women must fall into this category. To be sure, such a reaction would not be sensible for young adulthood, but is certainly a valid one for older ages. The purpose of this section is to show that this effect is far from being the only driver of adult excess female mortality.

To do so, we consider the number of missing women relative to female deaths in a given age group. As discussed in Section 2, this relative number is well captured by the ratio

$$\frac{d^w(a)/d^m(a)}{\hat{d}^w(a)/\hat{d}^m(a)}$$

which is always greater than 1 when there is excess female mortality but moreover, is larger the higher the "intensity" of excess

**Figure 2: Relative Excess Female Mortality at Different Ages**

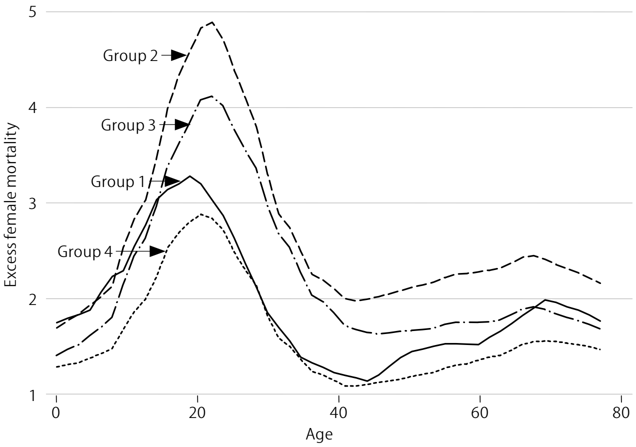
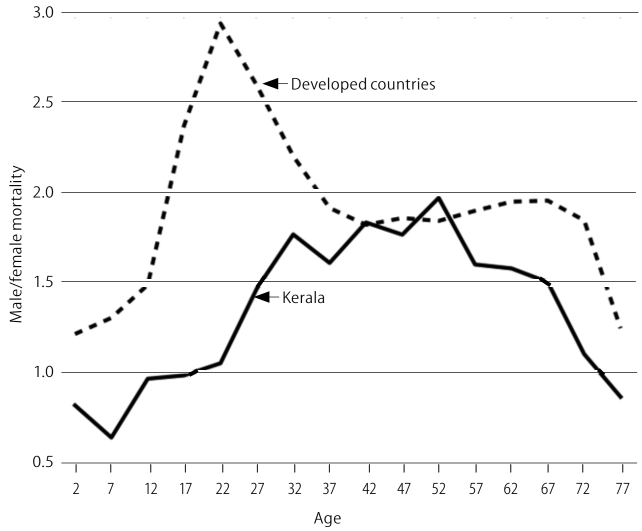



**Figure 3: Relative Mortality Rates: Kerala and Developed Countries**




female mortality at that age. Indeed, this ratio is the same as the actual death rate for women at age $a$ compared to the unbiased rate at that age, and so is comparable to $\sigma(0)/\hat{\sigma}(0)$, which is the corresponding ratio for the intensity of the problem at birth.

Figure 2 plots excess female mortality at different ages, along with relative sex ratios at birth. We group states into four categories. The first group (Group 1) comprises those states with excessively high sex ratios at birth (Punjab, Haryana). Group 2 includes those states with the highest proportion of the female population which is missing (refer to column 2 of Table 3). This group includes Madhya Pradesh, Maharashtra, Bihar and Assam, where the percentage of the female population which is missing is more than 0.65. Group 3 comprises those with a lower, but higher than average, proportion of missing women (greater than 0.50%). These include West Bengal, Orissa, Himachal Pradesh and Uttar Pradesh. A final group, Group 4, comprises those states with a lower proportion of missing women (less than or equal to 0.45%). This group includes Karnataka, Rajasthan, Gujarat, Andhra Pradesh, Tamil Nadu and Kerala.

We see from the figure that all groups exhibit excess female mortality at all ages. But relative excess female mortality is generally higher at adult ages, sharply highest at reproductive ages, with another gradual peak at old age. Adult excess female mortality is not just a reflection of higher adult mortality.

For Group 2, with the highest proportion of missing women overall, we see that relative to the other states, excess female mortality in these states is higher at all ages, except in childhood when it is on par with Group 1, the states with the very high sex ratios at birth. Group 3 has lower excess female mortality in childhood but higher excess female mortality in adulthood relative to Group 1. Those states with the lowest overall proportion of missing women (Group 4) have the lowest excess female mortality at all ages. However in adulthood, the levels of relative excess female mortality are not far off from those in Group 1. Overall, the differences in relative excess female mortality across the states are largest at adult ages (after age 15).

## 6 Kerala as a Benchmark

There is always the possibility that the use of developed countries as a reference group may be "inappropriate" for a poor country like India. Elsewhere (Anderson and Ray 2010) we have discussed this issue in some detail, and there is little we can add here at a conceptual level. Instead, we redo our computations using the state of Kerala as a benchmark.

As Table 4 makes clear, Kerala is the state in India with the lowest excess female mortality. That is not to say that there isn't any of it relative to developed countries. Figure 3 plots relative mortality rates at different age groups for Kerala and compares these to the corresponding rates for developed countries. It is clear that relative female mortality rates are still high, particularly at ages less than 45. (Indeed, the absence of any discrepancy would have rendered the discussion of reference groups moot.) Given this fact, our estimates of missing women by state will be reduced once Kerala is used as the reference, but our main results will still be shown to persist.

Table 5 describes total numbers of missing women by state with Kerala as reference. This table is analogous to Table 3. There are three main features of this table. First, the total number of missing women in the year 2003 is reduced by about a half,

**Table 5: Missing Women by Indian State (Kerala as Reference)** (2003, in 000s)

| State | Missing | % of Female Population | Proportion of Total |
|---|---|---|---|
| Haryana | 87 | 0.85 | 7.2 |
| Madhya Pradesh | 174 | 0.55 | 14.3 |
| Punjab | 62 | 0.57 | 5.1 |
| Maharashtra | 227 | 0.52 | 18.7 |
| Bihar | 177 | 0.42 | 14.6 |
| Assam | 50 | 0.43 | 4.1 |
| Uttar Pradesh | 159 | 0.18 | 13.1 |
| Himachal Pradesh | 11 | 0.36 | 0.9 |
| Orissa | 48 | 0.24 | 3.9 |
| West Bengal | 122 | 0.30 | 10.0 |
| Karnataka | 24 | 0.09 | 2.0 |
| Rajasthan | 34 | 0.13 | 2.8 |
| Gujarat | 18 | 0.08 | 1.5 |
| Andhra Pradesh | 18 | 0.05 | 1.5 |
| Tamil Nadu | 5 | 0.02 | 0.4 |
| Kerala | 0 | 0.00 | 0.0 |
| India | 1,219 | 0.23 | – |

Sources: National Family Health Survey (2005-06), Sample Registration System (Government of India).

which amounts to a million women. That is hardly surprising, as Kerala had missing women relative to the developed-country benchmark to begin with. Second, measured in terms of the proportion of the female population that is missing, the rankings across the states broadly persist, with the significant exception of Uttar Pradesh which performs much better under the Kerala benchmark. Third, in terms of state contributions to the total number of missing women, the relative burden shifts decisively towards the northern states. This has nothing to do with the resetting of Kerala's contribution to zero, as it was small to begin with anyway. It has more to do with the similarity across southern states in the age profile of relative mortality.

To see this and related points more clearly, Table 6 performs the same exercise as Table 4, but uses the four-group classification that we unearthed in the previous section.

The most important finding of Table 6 is that the pattern of missing women across age groups persists. Overall, the majority of missing women are still found at ages older than 15. One easy way to see this is to compare the all-India compositions across Tables 4 and 6. The use of Kerala as a benchmark has largely reduced the proportion of younger missing women (in the 0-15 age bracket). That is in line with Figure 3, which shows that Kerala exhibits significant excess female mortality in this very bracket relative to developed countries.

Finally, the four groupings still make sense. Group 1 still has the highest proportion of females missing at birth, while Group 2 has the high proportions of missing women overall, and these are distributed over all ages (post-partum), with the majority in the 15+ category. What is more, the use of Kerala as a benchmark reduces the number of missing women the least in these two groups. In contrast, the total number of missing

**Table 6: Missing Women by Indian State and Age Group (Kerala as Reference)**
(2003, in 000s)

| State | At Birth | | 0-15 | | 15-45 | | 45+ | |
|---|---|---|---|---|---|---|---|---|
| | Number | % | Number | % | Number | % | Number | % |
| Group 1 | | | | | | | | |
| Haryana | 44 | 50.0 | 20 | 23.1 | 6 | 7.3 | 17 | 18.9 |
| Punjab | 48 | 77.8 | 5 | 8.8 | 0 | 0 | 8 | 13.4 |
| Group 2 | | | | | | | | |
| Madhya Pradesh | 0 | 0 | 56 | 32.5 | 53 | 30.7 | 64 | 36.8 |
| Maharashtra | 15 | 6.7 | 34 | 14.9 | 31 | 13.7 | 147 | 64.6 |
| Bihar | 0 | 0 | 63 | 35.7 | 33 | 18.5 | 81 | 45.8 |
| Assam | 0 | 0 | 8 | 14.9 | 18 | 36.3 | 25 | 48.7 |
| Group 3 | | | | | | | | |
| West Bengal | 0 | 0 | 12 | 10.1 | 34 | 28.0 | 75 | 61.8 |
| Orissa | 9 | 18.6 | 0 | 0 | 13 | 27.7 | 26 | 53.7 |
| Himachal Pradesh | 1 | 6.0 | 2 | 17.7 | 2 | 17.9 | 7 | 58.3 |
| Uttar Pradesh | 0 | 0 | 14 | 8.8 | 80 | 50.3 | 65 | 40.8 |
| Group 4 | | | | | | | | |
| Karnataka | 0 | 0 | 0 | 0 | 10 | 42.5 | 14 | 57.5 |
| Rajasthan | 23 | 65.9 | 6 | 18.9 | 5 | 15.1 | 0 | 0 |
| Gujarat | 6 | 33.1 | 4 | 23.4 | 6 | 33.9 | 2 | 9.6 |
| Andhra Pradesh | 2 | 13.1 | 0 | 0 | 0 | 0 | 16 | 86.9 |
| Tamil Nadu | 0 | 0 | 0 | 0 | 1 | 30.5 | 3 | 69.5 |
| Kerala | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 149 | 12.2 | 226 | 18.6 | 295 | 24.2 | 550 | 45.1 |

Sources: National Family Health Survey (2005-06), Sample Registration System (Government of India).

women are reduced the most, by more than a factor of five, for the states in Group 4, with the lowest rates of excess female mortality under the earlier calculations. As we have mentioned before, not only do these states have low excess female mortality to begin with, like Kerala, but the age composition of relative mortality rates in these states also match that of Kerala.

The only significant change, is for two states in Group 4 (those states with the lowest rates of excess female mortality). The majority of missing women are now found at younger ages for Rajasthan and Gujarat. Relative male-to-female mortality patterns after birth (and especially in adulthood) in these states conform closely to that of Kerala. But they still have higher sex ratios at birth.

## 7 Channels

Because there is so much state-wise variation in the distribution of missing women across the age groups, it is difficult to provide a clear explanation for missing women in India. The traditional explanation for missing women, a strong preference for the birth of a son, is most likely driving a significant proportion of missing women in the two states of Punjab and Haryana where the biased sex ratios at birth are undeniable. However, the explanation for excess female deaths after birth is far from clear. It is beyond the scope of this paper to specifically identify the different pathways that explain the state-wise variation in the distribution of missing women across India. In what follows, we speculate on possible contributing factors.

We find significant excess female mortality in childhood everywhere except for the southern states (Kerala and Tamil Nadu have particularly low numbers of missing girls). Our earlier work (Anderson and Ray 2010) demonstrated that the bulk of missing women in India who die prematurely in childhood, do so when they are younger than five years old. We demonstrated that at least half of these excess deaths (1,21,000 per year) are from infectious and parasitic diseases. These mainly include vaccine preventable diseases and diarrhoea-related diseases, which are typically linked to poor sanitation conditions. It is difficult to disentangle gender discrimination from other factors such as biological, social, environmental, behavioural, or economic in explaining excess female deaths from these diseases. For example, it is conceivable that parents are more likely to give clean water to young boys compared to girls, but it is also possible that girls are inherently more susceptible to certain diseases. There is some evidence of gender discrepancies with regards to immunisation rates of young children (Pande and Yazbeck 2003). However, to conclude that this unequal treatment by parents explains all of the missing girls from vaccine-preventable diseases requires further research.

Two other factors that point more directly at gender discrimination and possibly to female infanticide, are the large number of excess female deaths under perinatal and congenital conditions in the childhood category. In our earlier work, the former accounted for 38,000 excess female deaths in a given year in India; the latter for over 13,000.

Consistent with our earlier work, we find significant excess female mortality at the reproductive ages, particularly in the central and north-eastern states. Our earlier work demonstrated that there were two key causes behind these excess deaths of women in India at this stage in their lives. The first is maternal mortality. If we compare maternal mortality rates and the percentage of the female population that is missing at reproductive ages across the states, we do indeed see a positive correlation.[14]

The other main cause of excess female mortality at the reproductive ages was identified in our earlier work to be "Injuries". In particular, in a given year, we estimated that excess female deaths for women in India from this cause exceeded 2,25,000, a number that dwarfed our maternal mortality estimates of 1,30,000 each year. These excess female deaths from "Injuries" would appear to be an indicator of overt violence against women. For the country as a whole, fire-related death is a leading cause; each year over 1,00,000 women are killed by fires alone. These excess deaths could well be associated with the custom of dowry which has been linked to bride-burning and dowry-death if promised dowry payments are not forthcoming.[15] Our estimates of excess female mortality at the reproductive ages are highest for the north-eastern region of India. It is indeed the case that the dowry phenomenon is more pervasive in northern parts of the country.[16]

Again consistent with our earlier work, the majority of missing women in India are found at older ages. Comparing the different states, excess female mortality in this age group is highest in the north-eastern states and lowest in the southern ones. Our earlier work demonstrated that excess burden falls mainly on noncommunicable diseases. Cardiovascular disease is particularly implicated. In India, women die at a rate closer to men from cardiovascular disease relative to developed countries. The plight of older women in the Indian subcontinent, especially of widows, has received some attention in the literature (see, e g, Drèze (1990), Chen and Drèze (1992) and Kochar (1999)). Our findings are fully supportive of this emphasis. Again, the extent to which excess female mortality in this age group is due directly to gender discrimination and differential access to health care is for future research.[17]

## 8 Conclusion

Our study of excess female deaths by age across the states of India yields two key findings. First, Indian women face the risk of excess mortality at every stage of their lives. Second, there is huge variation across the states in the distribution of missing women. Consistent with earlier findings, we find missing women at birth to be highest in two north-western states. Likewise we find excess female mortality at all ages to be lowest in several southern states. However, these regions make up less than a quarter of the total missing women in India. The bulk of missing women are distributed across the remaining states and die in adulthood. Amongst these states, there is variation as well. In some states, excess female mortality in childhood is on par with the north-western states, but excess female mortality in adulthood is significantly higher. For another set, excess female mortality in childhood is lower, but excess female mortality in adulthood is still very high.

Estimates of missing women were originally meant to represent some measure of the degree of gender discrimination. Our work demonstrates that much more research is needed to identify the underlying mechanisms that explain the missing women phenomenon in India. As discussed in Section 7, we have to go well beyond sex-selective abortion or female infanticide in our search for channels, at least as far as India is concerned. Sen's "terrible story of inequality and neglect" is possibly true in large part, but other stories may need to be told as well. In our view, an entire gamut of interpretations is possible. For example, gender-based violence and inadequate maternal healthcare may explain a significant proportion of missing women at the reproductive ages. These are clear and overt signs of "inequality and neglect". At older ages, excess female deaths may stem from "unequal treatment", but the notion needs to be amplified.

A good example is the cardiovascular deficit which presents another sort of interpretative quandary. We have already remarked that heart disease accounts for a large fraction of excess female mortality.[18] There is an entire array of hypotheses to explain the phenomenon. It could be genetic: for instance, the recently discovered "heart disease gene" so prevalent in south Asia may be equally present in males and females, thereby lowering the gender skew in incidence. Lifestyle differences by gender may be important: diet, attention to personal health and well-being, and so on. Or it may truly be lack of overt "similar care": women seek or receive medical care less often in developing countries, or may be subject to greater stress. Put another way, if we want to restrict ourselves to defining missing women as the number of females who have died due to overt discrimination, then the original estimates need to be seriously revised downwards. Moreover, such a computation is not at all straightforward.

More importantly, however, what is most clear from our exercise is that the plight of adult women in India is as serious a problem as that of young girls who were either never born or die prematurely in childhood. That raises a final philosophical point. Is there reason to value the "excess death" of an adult any less than that of a child, or one never born? The answer is perhaps easier in the comparison of an infant and an elderly person: consider all the years of life that were never lived for the former. Between infants and young women, however, the comparison is far less clear. But these important philosophical considerations are the subject of another paper.

## NOTES

1 Any computation of missing women presupposes a counterfactual. For Sen this counterfactual is just the overall sex ratio in countries where men and women presumably "receive similar care". Sen's baseline ratio was the average overall sex ratio for developed countries today. Others have revised these estimates using alternative benchmarks (Coale 1991 and Klasen and Wink 2003), which still yield enormous figures.

2 Jha et al (2006) also observe that in India, the ratios of males to females for second- and third-order births, conditional on the previous births being female, are extremely high.

3 See, e g, Deaton (1989); Subramanian and Deaton (1991); Garg and Morduch (1998). There is some evidence that excessive female mortality at these younger ages falls most heavily on girls at higher birth orders (see, for example, Das Gupta 1987). Gender-based stopping rules may also contribute to differential mortality rates, as girls are likely to be members of larger (and therefore, ceteris paribus, poorer and somewhat more death-prone) families.

4 See Bhaskar and Gupta (2007); Klasen and Wink (2003); Mari Bhat (2002a, 2002b).

5 Drèze (1990); Chen and Drèze (1992); Kochar (1999); Mari Bhat (2002b, 2002c); Jayaraj (2009) discuss female mortality at older ages.

6 We use the group of Established Market Economies as defined by the World Bank: Western Europe, Canada, United States, Australia, New Zealand, and Japan.

7 We exclude Jammu and Kashmir from the analysis due to lack of data.

8 See also Dubuc and Coleman (2007) for evidence from the UK.

9 We do not know how true this might be for India or south Asia, though we certainly know from the work of Chahnazarian (1988) and others that there are systematic differences for African populations. There are surely additional corrections that are possible. For instance, there is some evidence that environmental factors determine the SRB (see Anderson and Ray (2010) for further discussion of this issue).

10 This is the only publicly available data on sex ratios at birth by year and state.

11 See, for example, Jayaraj and Subramanian (2004).

12 This number is obtained by doing the calculations for India as a whole, and not by adding up the numbers across states. It can be shown that the "missing women" formula is not, in general, linearly decomposable across subunits.

13 Gujarat could potentially be included in "conventional-wisdom" category since the total missing women under age 15 add up to 50% of missing women in the state. However, as pointed out in Section 3, the number of missing women at birth in Gujarat is likely overestimated.

14 Maternal mortality rates by state come from "Maternal Mortality in India 1997-2003: Trends, Causes, and Risk Factors" Sample Registration System of India (2006).

15 See, for example, Bloch and Rao (2002). The National Crime Bureau of the Government of India reports approximately 6,000 dowry deaths every year, but numerous incidents of dowry-related violence are never reported. Menski (1998) puts the number at roughly 25,000 brides who are harmed or killed each year. Our indirect estimates could make these numbers much higher.

16 See Dalmia and Lawrence (2010).

17 Refer to Karkal (1999) for a discussion of gender discrimination and aging in India.

18 See Anderson and Ray (2010) and the references therein.

## REFERENCES


Abrevaya, J (2009): "Are There Missing Girls in the United States? Evidence from Birth Data", *American Economic Journal: Applied Economics*, 1, 1-34.

Almond, D and L Edlund (2008): "Son-Biased Sex Ratios in the 2000 United States Census", *Proceedings of the National Academy of Sciences*, 105(15), 5681-82.

Anderson, S and D Ray (2010): "Missing Women: Age and Disease", *Review of Economic Studies*, 77, 1262-1300.

Bhaskar, V and B Gupta (2007): "India's Missing Girls: Biology, Customs, and Economic Development", *Oxford Review of Economic Policy*, 23(2), 221-38.

Bloch, F and V Rao (2002): "Terror as a Bargaining Instrument: Dowry Violence in Rural India", *American Economic Review*, 92 (4), 1029-43.

Chahnazarian, A (1988): "Determinants of the Sex Ratio at Birth: Review of Recent Literature", *Social Biology*, 35(3-4), 214-35.

Chen, M and J Drèze (1992): "Widows and Well-Being in Rural North India", mimeo, London School of Economics.

Coale, A (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*, 17(3), 517-23.

Dalmia, S and P G Lawrence (2010): "Dowry Inflation in India: An Examination of the Evidence for and against", *Journal of International Business and Economics*, 10(3), 58-74.

Das Gupta, M (1987): "Selective Discrimination against Female Children in Rural Punjab, India", *Population and Development Review*, 13(1), 77-100.

– (2005): "Explaining Asia's 'Missing Women': A New Look at the Data", *Population and Development Review*, 31(3), 529-35.

Deaton, A (1989): "Looking for Boy – Girl Discrimination in Household Expenditure Data", *World Bank Economic Review*, 3, 183-210.

Drèze, J (1990): "Widows in Rural India", mimeo, London School of Economics.

Dubuc, S and D Coleman (2007): "An Increase in the Sex Ratio of Births to India-Born Mothers in England and Wales: Evidence for Sex-Selective Abortion", *Population and Development Review*, 33, 383-400.

Garg, A and J Morduch (1998): "Sibling Rivalry and the Gender Gap: Evidence from Child Health Outcomes in Ghana", *Journal of Population Economics*, 11(4), 471-93.

Hill, K (1987): "Estimating Census and Death Registration Completeness", *Asian and Pacific Population Forum*, 1(3): 8-13, 23-4.

Jayaraj, D (2009): "Exploring the Importance of Excess Female Mortality and Discrimination in 'Natality' in Explaining the 'Lowness' of the Sex Ratio in India", *The Developing Economies*, 47(2), 177-201.

Jayaraj, D and S Subramanian (2004): "Women's Well-Being and the Sex Ratio at Birth: Some Suggestive Evidence from India", *Journal of Development Studies*, 40, 91-119.

Jha, P, R Kumar, P Vasa, N Dhingra, D Thiruchelvam and R Moineddin (2006): "Low Male-to-Female Sex Ratio of Children Born in India: National Survey of 1.1 Million Households", *Lancet*, 367, 211-18.

Karkal, M (1999): "Ageing and Women in India", *Economic & Political Weekly*, 34(44), WS54-WS56.

Klasen, S and C Wink (2003): "'Missing Women': Revisiting the Debate", *Feminist Economics*, 9(2-3), 263-99.

Kochar, A (1999): "Evaluating Familial Support for the Elderly: The Intrahousehold Allocation of Medical Expenditures in Rural Pakistan", *Economic Development and Cultural Change*, 47, 620-56.

Mari Bhat, P N (2002a): "Completeness of India's Sample Registration System: An Assessment Using the General Growth Balance Method", *Population Studies*, 56(2), 119-34.

– (2002b): "On the trail of 'Missing' Indian Females: I: Search for Clues", *Economic & Political Weekly*, 37(51), 5105-18.

– (2002c): "On the Trail of 'Missing' Indian Females: II: Illusion and Reality", *Economic & Political Weekly*, 37(52), 5244-63.

Mathers, C, A Lopez and C Murray (2006): "The Burden of Disease and Mortality by Condition: Data, Methods, and Results for 2001" in A Lopez, C Mathers, M Ezzati, D Jamison and C Murray (ed.), *Global Burden of Disease and Risk Factors*, Oxford University Press and The World Bank, pp 45-240.

Menski, W, ed. (1998): *South Asians and the Dowry Problem* (New Delhi: Vistaar Publications).

Pande, R P and A S Yazbeck (2003): "What's a Country Average? Wealth, Gender and Regional Inequalities in Immunisation in India", *Social Science and Medicine*, 57, 2075-88.

Sen, A (1990): "More Than 100 Million Women Are Missing", *The New York Review of Books*, 37(20), 20 December.

– (1992): "Missing Women", *British Medical Journal*, 304, March, 587-88.

Subramanian, S and A Deaton (1991): "Gender Effects in Indian Consumption Patterns", *Sarvekshana* 14, 1-12.


## Data Appendix

For mortality rate estimates for the year 2003 from India, we use data provided by the Sample Registration System (SRS) of India. We use the average mortality rates across the years (2001-06). Adjusting the margins of this aggregation of data made negligible changes to our estimates. For example, we alternatively tried an average from the years (2000-06), or just used data from the year 2003. To address under-reporting in the data, following the procedures used by the WHO, we use the Generalized Growth Balance Method (refer to Mathers et al (2006) and to Mari Bhat (2002a)). This methodology estimates the completeness of death recording relative to population recording. It uses two periods of population estimates (here we use 2001 and 2006) and inter-year death rate estimates (we use the year 2003) to estimate the degree of completeness of the death data. Refer to Hill (1987) for more details regarding this technique. Using this method, we estimate that, for India, about 12% of deaths are under-reported in the year 2003, and that the degree of under-reporting actually increases with age: where for ages 0-15, under-reporting is at roughly 7%, whereas for ages (15-45) it is closer to 12% and for ages older than 45, it is about 13%. We also find that the under-reporting of deaths is significantly higher for females than for males, in particular, for women aged 15 and older, the under-reporting of deaths is closer to 20%. There is also variation across the states, where, for example, the under-reporting of deaths is lowest in Kerala (this was similarly found in Mari Bhat (2002a) using earlier years of SRS data).