THE REVIEW OF ECONOMIC STUDIES

Review of Economic Studies (2010) **77,** 1262–1300
0034-6527/10/00411011$02.00

doi: 10.1111/j.1467-937X.2010.00609.x

Missing Women: Age and Disease

SIWAN ANDERSON
University of British Columbia

and

DEBRAJ RAY
New York University

First version received July 2008; *final version accepted December* 2009 *(Eds.)*

Relative to developed countries and some parts of the developing world, most notably sub-Saharan Africa, there are far fewer women than men in India and China. It has been argued that as many as a 100 million women could be missing. The possibility of gender bias at birth and the mistreatment of young girls are widely regarded as key explanations. We provide a decomposition of these missing women by age and cause of death. While we do not dispute the existence of severe gender bias at young ages, our computations yield some striking new findings: (1) the vast majority of missing women in India and a significant proportion of those in China are of adult age; (2) as a proportion of the total female population, the number of missing women is largest in sub-Saharan Africa, and the absolute numbers are comparable to those for India and China; (3) almost all the missing women stem from disease-by-disease comparisons and not from the changing composition of disease, as described by the epidemiological transition. Finally, using historical data, we argue that a comparable proportion of women was missing at the start of the 20th century in the United States, just as they are in India, China, and sub-Saharan Africa today.

1. INTRODUCTION

The phrase "missing women", coined 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. On average, males outnumber females at birth, but that imbalance begins to redress itself soon after.[1] The combined effect is (or should be) a roughly equal proportion of men and women in the population as a whole.[2] That is not the case in large parts of Asia: in India and China, the overall ratio of males to females is around 1.06.

By contrast, sub-Saharan Africa appears to do remarkably well. Sen (1990) observes that this region, "ravaged as it is by extreme poverty, hunger, and famine, has a substantial excess rather than deficit of women", and conjectures that the high female participation in the labour force there plays a role in "linking women's gainful employment and survival prospects".

1. See, e.g. Coale (1991) and the Regional Model Life Tables in Coale, Demeny, and Vaughan (1983).
2. The ratio varies a bit depending on life expectancy and population growth; see Coale (1991). With "moderate" life expectancies of 60 for women and 56.5 for men, and using the West model life tables, the ratio slightly favours men (1.022) when population is growing at 2% per year, and women (0.997) in stationary populations.

Much subsequent attention has been placed on a skewed sex ratio at birth, which may indicate sex-selective abortion.[3] For instance, Zeng *et al*. (1993) and Das Gupta (2005) observe that sex ratios for higher order births in China (conditioning on earlier births being female) are significantly skewed towards males, a clear warning sign of sex selection through abortion or infanticide.[4] A second area of focus is early childhood and the possibility that young girls are systematically less cared for.[5] It is not that the later-age travails of women are entirely ignored (see, for instance, Drèze, 1990; Chen and Drèze, 1992; Kochar, 1999), but the accusing finger seems to have been predominantly pointed at the prenatal and infant/early childhood stages. Das Gupta (2005) quite fairly summarizes the literature when she states that "the evidence indicates that parental preferences overwhelmingly shape the female deficit in South and East Asia".

This paper performs an elementary accounting exercise. We study how missing women in three different regions—India, China, and sub-Saharan Africa—are distributed across all age groups (and then by age-disease groups). The methodology we employ is in the spirit of the Sen contribution. For each category—age, and later, age and disease—we posit a "reference" death rate for females, one that would be obtained if the death rate of males in that country were to be rescaled by the *relative* death rates for males and females (in the same category) in developed countries.[6] We subtract this reference rate from the actual death rate for females, and then multiply by the population of females in that category. This is the definition of "missing women" in the category of interest. Whether it corresponds to some intuitive notion of excess female mortality due to lack of similar care across gender is a complex issue, one that we return to at several points in this paper.

We study the distribution of missing women by age and disease, as well as by the aggregate numbers. Our age decomposition yields some striking findings. First, although the *overall* sex ratios in India and China are similar—both around 1.06—the two countries have distinct age profiles of missing women. A large percentage of the missing women in China are located before birth and in infancy. We estimate that around 37–45% of the missing women in China are due to prenatal factors alone. But the numbers for India are more evenly distributed across the different age groups. Prenatal factors account for around 11%, and if we add up all the female deficit up to age 15, we do not get to a third of the total.

Second, we find that, as a proportion of the female population the annual flow of missing women is actually largest in sub-Saharan Africa, *despite the fact that their overall male to female ratio is less than unity*. None of this happens at birth; missing women are spread over the entire age spectrum. When we add up the flow numbers by age for the year 2000, around 1.53 million women are missing in sub-Saharan Africa for that year alone, which is not that different from the Chinese and Indian estimates of 1.73 and 1.71 million, respectively. Expressed as a fraction of the female population, the sub-Saharan numbers are significantly higher than their Chinese and Indian counterparts.

These findings can be reconciled with the balanced overall sex ratio in sub-Saharan Africa by the relatively low sex ratio at birth there. In sub-Saharan Africa, the sex ratio at birth is

3. See, e.g. Junhong (2001); Sudha and Rajan (1999); Ebenstein (2007) and Lin *et al*. (2007).

4. Similarly, Jha *et al*. (2006) 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.

5. See, e.g. Deaton (1989); Subramanian and Deaton (1991); Garg and Morduch (1998), and Oster (2009). There is some evidence that excessive female mortality at these younger ages falls most heavily on girls at higher birth orders (see, e.g. 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.

6. These are the so-called Established Market Economies as defined by the World Bank: Western Europe, Canada, United States, Australia, New Zealand, and Japan.

approximately 1.03, much lower than for developed countries as a whole (which is 1.06). There appear to be genetic differences that determine this ratio.[7]

We then turn to decompositions by age and disease. With birth set aside, at younger ages the bulk of missing females come primarily from infectious diseases. At older ages, the majority of them can be traced to non-communicable diseases. At the same time, there are systematic differences across the three regions.

In India, respiratory and infectious diseases are important sources of excess female deaths, of the same order of magnitude as maternal mortality. But two other categories stand out. The main cause (by a long way) of excess female deaths in India is cardiovascular disease. It dominates all other sources of excess female mortality, and easily outstrips missing females at birth. The second and rather ominous category for excess female deaths in India is "Injuries". There are more missing women here, for instance, than the total from maternal mortality.

In sub-Saharan Africa, much of the female deficit is to be found at younger ages. Malaria is an important component, and so is maternal mortality: much more so than in India. But *the* dominant source of missing women is the human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS). We estimate that there are over 600,000 excess female deaths each year from this source alone. That said, it is still true that the percentage of missing women in sub-Saharan Africa is comparable to that in India or China even if HIV/AIDS is ignored. For instance, the female deficit in cardiovascular disease is also large in sub-Saharan Africa.

In China, infectious diseases do not play a large a role. The main female deficit is at older ages where non-communicable diseases, such as cardiovascular and respiratory diseases, are important.

A detailed description of these findings is postponed to the main sections of the paper, but one question arises immediately. Is it possible that the difference in mortality rates by gender between developing and developed countries is simply due to changing compositions of disease? In this case, excess female deaths in developing countries would arise not from lack of "similar care" for men and women, but from the changing disease environment. The epidemiological transition suggests this might be the case. That transition refers to changes in the causal composition of mortality (with development), with infectious diseases giving way to chronic and degenerative ailments as the leading causes of death. Conventional wisdom states that infectious diseases and undernutrition do not discriminate across gender, while there is a female advantage in chronic and degenerative disease. Age-by-age calculations that do not control for compositional effects of disease composition might then record a number of "missing women", but they would not be missing due to any lack of "similar care".

Our decomposition allows us to address this issue. Contrary to what we might expect, we find that the epidemiological transition plays a minimal role.[8] Instead, disease-by-disease effects account for an overwhelming fraction of all missing women. In India, the number is 1.64 million out of the aggregate estimate of 1.71 million, in sub-Saharan Africa 1.39 million out of a total of 1.53 million, and in China 1.59 million out of 1.73 million.

These results compel us to confront the question of various ages and various diseases when studying missing women. The aggregate female deficit in South and East Asia has been mainly attributed to parental preferences which discriminate against young or unborn girls. Our findings suggest that excess female mortality is a more universal phenomenon (both over age and over disease).

7. Indeed, the sex ratio at birth for whites in the United States is around 1.06 whereas for blacks it is 1.03. The available data for births from sub-Saharan parents *in the United States* suggests similar numbers as well.

8. We are using the term "epidemiological transition" loosely here. We simply mean the effect of a changing disease mix across countries.

We take our methodology to historical data for now-developed countries. We compute the number of missing women in the United States in 1900. There is a remarkable congruence between these numbers and what we observe in the three regions today. Expressed as a proportion of the female population, the number of missing women in the United States in 1900 is *larger* than in India and China today, and slightly smaller than in sub-Saharan Africa.

These results raise the spectre of interpretation. What constitutes a valid benchmark against which to measure missing women? We use the relative death rates of males and females in developed countries today, but perhaps there is "natural" variation in such relative rates with development. Our position, on which we expand in Section 6, is that there is little or no reason for such variation and, moreover, that an examination of existing data for poor countries cannot settle the issue because we have no *separate* way to argue that those countries exhibit no gender discrimination. That includes the historical United States; we have no reason to believe that it was discrimination-free.[9]

At the same time, our findings raise questions about just what "missing women" means. Section 6 addresses this issue in detail. In our view, an entire gamut of interpretations is possible. For some categories, such as missing women from Injuries, we would argue the dominant cause must be discrimination across gender. On others, such as missing women from HIV/AIDS, the situation is more complex. There may be lack of similar care in treatment, there may be gender-based violence, and some of the excess may also be due to different cultural and sexual norms. Similarly, in the case of cardiovascular illness, excess female deaths may stem from unequal treatment, but also may be due to differential incidence. Put another way, if we want to restrict ourselves to defining missing women as the number of females who have died due to discrimination, *then the original estimates need to be seriously revised downwards*. Moreover, such a computation is not at all straightforward.

An accounting exercise cannot separate the role of direct gender discrimination from other factors—biological, social, environmental, behavioural, or economic—in explaining excess female mortality. But it allows us to ask these questions, and to assess the comparative importance of each potential category. We may (or may not) already know that the sex ratio at birth is skewed, that there is a female deficit from cardiovascular disease, that women die relatively more from injuries in India, that there is maternal mortality, and that HIV/AIDS disproportionately kills young women in sub-Saharan Africa. Our decomposition puts all these varied sources into one unified and *comparable* framework, and it therefore has the potential to inform future research in these areas.

2. COMPUTING MISSING WOMEN

2.1. *The Sen counterfactual*

Sen describes how skewed sex ratios can be translated into absolute numbers of missing women:

> To get an idea of the numbers of people involved in the different ratios of women to men, we can estimate the number of "missing women" in a country, say, China or India, by calculating the number of extra women who would have been in China or India if these countries had the same ratio of women to men as obtains in areas of the world in which

9. Indeed, we cannot be sure about lack of discrimination in *any* country. All we can argue is that the developed countries today set a norm that other countries, including the historical United States, can be measured against.

> they receive similar care . . . . In China alone this amounts to 50 million "missing women" . . . . When that number is added to those in South Asia, West Asia, and North Africa, a great many more than 100 million women are "missing". These numbers tell us, quietly, a terrible story of inequality and neglect leading to the excess mortality of women. (Sen, 1990)

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". True, Sen's baseline ratio—the average overall sex ratio for Europe, North America, and Japan—is somewhat optimistic for female survival, including as it does war losses and a different age composition, but the more conservative numbers, most notably the alternative calculations by Coale (1991), still yield enormous figures: around 60 million.[10]

2.2. *Our approach*

Throughout, we apply variants of the Sen–Coale counterfactual, first to every age group, and then to age/disease groups. Briefly, we suppose (for each age or age-disease category) that the relative death rates of females to males are "free of bias" in developed countries. We compare these rates with the *actual* relative rates in the country of interest, and obtain missing women under that category.

What we do is no more and no less than a careful accounting exercise. We take the counterfactual seriously and see what it delivers. In particular, we address which age and disease categories house the missing women that are identified overall by Sen and Coale.

Two points are to be noted, and we return to them in Section 6. First, the choice of any counterfactual, including the one we use, can be challenged. We simply do not know if the "natural", "discrimination-free" relative death rates for women and men are the same in poor and in rich countries. To know this, we must somehow assert that some poor regions do *not* have discrimination and use the relative death rates for those supposedly discrimination-free regions as benchmarks instead. Such an alternative may be useful from the viewpoint of robustness, though far from conclusive.

Second, any missing women we do estimate may be "missing" for a variety of reasons. We do not suggest that all these numbers *must* be attributed to, say, discrimination. Indeed, our discussion throughout and especially in Section 6 will explicitly assert that there are several potential channels, all worth exploring in future research. Sen's "terrible story of inequality and neglect" is possibly true in large part, but other stories may need to be told as well.

The fact that there may be excess female deaths at various age groups is well known, and we do not claim to have discovered this.[11] Our contribution is two-fold. First, we develop a precise decomposition *and* take it to the data for developing countries, thereby obtaining a breakdown of missing women by age in India, China, and sub-Saharan Africa. To our knowledge, these relative numbers by age have never been calculated or discussed before. Second, we take exactly the same idea to disease. Once more, it is surely the case that demographers are aware

10. Coale (1991, Table 1) uses the West model life tables to predict the overall sex ratios "that would exist in the absence of . . . discrimination". This leads to a lower estimate of missing women compared to Sen's, because—in Coale's words—the actual sex ratio in these countries "is an inappropriate standard: it is the result of past male war losses and of an age composition that reflects past low fertility", in addition to any absence of discrimination.

11. For instance, Klasen (1998) studies excess female mortality in Germany between 1740 and 1860 at two different age groups.

of differential gender mortality patterns across causes of death. Yet, we have never seen a unified presentation of the numbers that is comparable across age and disease groups.

2.3. *Missing women by age*

Begin with age-specific computations. Let a stand for an age group, where $a = 1, \ldots, n$. The extra value $a = 0$ denotes birth. For any age $a \geq 1$, deaths within that group a will refer 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 in the country or region of interest. Use the label $\widehat{\ }$ to denote these variables for the benchmark or reference region.[12] The *reference death rate* for women of age a in the country of interest is defined by

$$u^w(a) = \frac{d^m(a)}{\widehat{d}^m(a)/\widehat{d}^w(a)}. \tag{1}$$

The number of extra female deaths, and hence missing women, in the country of interest at age a in a given period is then equal to the difference between the actual and reference death rates for women, weighted by the number of women in that age group:

$$\mathrm{mw}(a) = \left[d^w(a) - u^w(a)\right]\pi^w(a), \tag{2}$$

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

2.4. *Missing women by age and disease*

We employ an entirely parallel calculation for missing women by age *and* disease. Consider any age $a \geq 1$, and denote by $d^m(a,k)$ and $d^w(a,k)$ the rates of death of men and women, respectively, from disease k at age a in the country of interest. The reference death rate of women at age a from disease k in the country of interest is then defined by

$$u^w(a,k) = \frac{d^m(a,k)}{\widehat{d}^m(a,k)/\widehat{d}^w(a,k)}. \tag{3}$$

The number of extra female deaths in the country of interest at age a from disease k in a given period is therefore equal to

$$\mathrm{mw}(a,k) = \left[d^w(a,k) - u^w(a,k)\right]\pi^w(a) \tag{4}$$

where $\pi^w(a)$, as before, is the starting population of women of age a.[13]

2.5. *Sex ratios at birth*

To the estimates in the previous sections we must now add the number of missing women at birth. This necessitates a choice of some "unbiased" reference sex ratio at birth, which is an extremely difficult question. Ideally, we want as a comparison point the sex ratio at birth generated by "the same group in the same circumstances", *minus* any differential treatment for boys and girls. Such a reference point is not available.

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

13. In Section 4.1, we modify this approach for certain causes of death, such as maternal mortality.

TABLE 1

Sex ratios at birth by nationality/ethnicity in the United States

Nationality/ethnicity	Sex ratio at birth
White	1.054
Black	1.030
Sub-Saharan African	1.035
Chinese	1.074
Asian Indian	1.066
American Indian	1.031
Japanese	1.055
Hawaiian	1.054
Filipino	1.072
Puerto Rican	1.045
Cuban	1.054
Central and South American	1.044
Mexican	1.041

Notes: The data on sex ratios at birth for all race/ethnicities groups (except for Asian Indian and sub-Saharan African) come from the National Vital Statistics of the United States. The averages reported in the table are a computation for the years 1970–2002. They do not vary substantially from just the most recent estimates for the year 2002, with the exception of Japanese who have a sex ratio at birth of 1.089 in that year. Data on the sex ratio at births for Asian Indians is not available at the national level before 1992; the estimate in the table is from Abrevaya (2009) for the years 1992–2004. The numbers for sub-Saharan African parents come from IPUMS United States, 2000.

Coale (1991) used a reference sex ratio at birth of 1.059 for all groups. This is problematic: there is substantial variation in the sex ratio at birth across race and ethnicity. The average sex ratio at birth for developed countries is in the range of 1.05 to 1.07, with a median equal to 1.059, but this range is non-trivial. Even within Europe, the average in Northern Europe is around 1.05, whereas for the Mediteranean it is in the range of 1.06 to 1.07.

More to the point, the sex ratio at birth is *significantly* shifted downward for African-American parents (see, e.g. James, 1987), and it is around 1.03. The available evidence suggests that this is also true of sub-Saharan African parents in the United States. In contrast, the sex ratios at birth for Asian populations (Asian Indians, Chinese, and Filipinos) in the United States is around 1.07. Table 1 summarizes some of this information.

Certainly, there are a number of behavioural, biological, and environmental factors which can explain part of the variation in sex ratios at birth.[14] These factors, however, do not explain the racial differences. In the United States, the lower sex ratio at birth for blacks and native populations compared to the white population has been observed for a long time and this large systematic variation found across ethnic/racial groups has persisted. Indeed, the sex ratio at birth for blacks and whites in the United States has remained relatively constant for at least a century.[15] Studies demonstrate that the strong racial effect persists when controlling for other

14. Biological determinants of the sex ratio at birth include the timing of conception and hormonal variations (James, 1987). However, these factors have proved difficult to measure and most research has relied on variables which are more easily observable at a large scale such as parental age and birth order. In general, the proportion of male births increases with the number of prior births and shorter birth intervals and it decreases with parental age and the proportion of multiple births.

15. The mean sex ratio at birth between 1915 and 1948 is 1.059 for whites and 1.029 for blacks (McMahan, 1951); between 1942 and 1963, they are 1.057 and 1.023, respectively (Tarver and Lee, 1968); and for 1970–2002, the respective averages are 1.054 and 1.030 (Mathews and Hamilton, 2005).

factors such as parental age, birth order, and parity and these latter effects, in turn, decrease in their importance (Chahnazarian, 1988).[16]

As for the higher ratios in Asian populations, darker forces may be at work. There is emerging evidence that the Asian Indian, Chinese, and South Korean populations residing in the United States may well be practising gender selection at the prenatal stage, like some of their fellow nationals back home (Almond and Edlund, 2008; Abrevaya, 2009).[17] However, it is unclear at this stage whether the phenomenon is pervasive enough to alter the overall estimates. For example, the estimated sex ratio at birth for the Chinese population residing in the United States between 1931 and 1936 is also around 1.07 and this estimate certainly predates access to ultrasound techniques. Similarly, the average sex ratio at birth among the Filipino population residing in the United States is in the same range as the other Asian groups. This community has typically not been associated with sex-selective practices.

It should also be pointed out that a sex ratio at birth 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 (particularly between blacks and whites), we should certainly be wary of using the sex ratio at birth of Asian populations residing in developed countries as a reference, and in what follows we use a range of numbers.

For India, we use a reference ratio between 1.059 and 1.066. The lower ratio is the average across developed countries and is the one used by Coale (1991), while the higher ratio is the average sex ratio at birth among Asian Indians in the United States (as in Table 1). For sub-Saharan Africa, we use the range 1.030–1.035. The lower end of the range is the well-documented sex ratio at birth for African-Americans. The upper end is a 2000 estimate using IPUMS data for sub-Saharan Africans in the United States (Table 1).[18] For China, we use a reference ratio in the range 1.059–1.074. As in the case of India, the lower bound is taken from the average across all developed countries, while the upper bound is drawn from Chinese populations in the United States (Table 1).

Turn now to actual sex ratios at birth in the regions of interest. The most recent estimates of the sex ratio at birth for Indians *in* India range from 1.070 to 1.078.[19] Those for China are substantially higher. The 2000 census places this ratio at 1.169. The UN Demographic Yearbook reports a lower number for 1989 (1.139). These large ratios are to be contrasted with those for sub-Saharan Africa. Using 56 Demographic Health and World Fertility Surveys that cover 29 sub-Saharan African countries, and comparing these with other studies (including

16. Some research has aimed to better understand this racial effect by examining the sex ratio at birth for interracial couples. These studies have concluded it is the father's race which matters. That is, white fathers coupled with either black or American Indian mothers still produced a higher proportion of male births, whereas white mothers did not (Khoury *et al*., 1984). Similar results were found for Korean fathers who formed interracial unions (Morton *et al*., 1967).

17. See also Dubuc and Coleman (2007) for evidence from the United Kingdom

18. In the light of footnote 20 below, this estimate probably needs to be broken up across individuals of Bantu and non-Bantu origin, though we doubt that this will make any difference to the analysis to follow.

19. The data come from the Demographic Health Survey (DHS, alternatively named the National Family Health Survey). Estimates for the sex ratio at birth include children born between 1980 and 1999. The more recent cohorts have lower sex ratios at birth. Between 1980 and 1990, the estimated ratio is 1.078, and between 1990 and 1999 it is 1.073. According to the most recent DHS (2005–2006), the estimated sex ratio at birth for the year 2000 is 1.070 and the average sex ratio at birth between 1996 and 2006 is 1.074. Both within India and China, there is substantial regional variation in this sex ratio at birth. For example, some of the northern states in India have particularly high sex ratios at birth (at least 1.10). But the all-India average is not in this range.

birth registration) where available, Garenne (2002, 2004) places sub-Saharan Africa as a whole at around 1.033. But there is substantial variation within the region.[20]

To compute missing women at birth, we compare the sex ratios at birth for the same group (Indians, sub-Saharan Africans, Chinese) with our best guess for the appropriate reference ratio. We use a formula analogous to (2) to carry out this computation:

$$\mathrm{mw}(0) = \left[\frac{\sigma(0)}{\widehat{\sigma}(0)} - 1\right]\pi^{w}(0) \tag{5}$$

where $\sigma(0)$ is the sex ratio at birth in our country, $\widehat{\sigma}(0)$ the comparison ratio from developed countries, and $\pi^{w}(0)$ is the total number of female births for the very same group.

There are surely additional corrections that are possible. For instance, there is some evidence that environmental factors determine the sex ratio at birth.[21] It is also likely that the sex ratio at birth is affected by development.[22] But it is unclear whether further fine-tuning would be that useful in the absence of credible information regarding the sex ratio at birth for these groups in developed countries. We therefore stick with the comparisons that we have. However, it may be useful to note that an additional 0.01 difference in the actual and comparison ratios means a difference of approximately 120,000 missing females (per year) in India and 90,000 missing females in China. These numbers are significantly less than 7% of the total number of missing women that we later estimate for each of these regions.

2.6. *Aggregation*

Our procedure allows us to generate a first estimate for missing women, one that controls for age composition as well as group-specific differences in the sex ratio at birth. This estimate, which we call mw_A, is given by

$$\mathrm{mw}_A = \sum_{a=0}^{n} \mathrm{mw}(a). \tag{6}$$

It is important to note that mw_A includes all changes in the disease composition as we compare across the region of interest and developed countries. But we can generate a second estimate for missing women that effectively keeps the disease mix unchanged: one that, in effect, controls for both the age composition and the disease mix. This estimate is obtained by simply adding missing women by age *and* disease over all ages and diseases:

$$\mathrm{mw}_B = \mathrm{mw}(0) + \sum_{a=1}^{n}\sum_{k} \mathrm{mw}(a,k). \tag{7}$$

20. Garenne argues that the predominantly Bantu populations of Eastern and Southern Africa exhibit sex ratios at birth below 1.000, while Nigeria and Ethiopia display high, Asia-like ratios. Finally, a large group of countries such as Ghana, Mali, and Côte d'Ivoire appear to be around the 1.050 mark.

21. For example, lower proportions of male offspring have been observed in populations exposed to dioxin, mercury, pesticides, polychlorinated biphenyls (PCBs) and also parental smoking (MacKenzie, Lockridge and Keith, 2005). Others have connected variations in the sex ratio at birth to wars Myers (1947), seasons (Lerchl, 1998), and latitude (Navara, 2009). There has been a recent debate on the effect of Hepatitis B on sex ratios (Oster, 2005; Lin and Luoh, 2008; Oster *et al*., 2008)

22. There is evidence that male offspring are more susceptible to death *in utero* than the female, so maternal malnutrition could be linked to a lower sex ratio at birth (Andersson and Bergstrom, 1998). With better health care, it is to be expected that the sex ratio at birth will rise as more male fetuses survive. Klasen and Wink (2003) attempt to correct for this; some reflection immediately shows that such a correction must *increase* the number of missing women at birth.

The second procedure deliberately makes no attempt to account for any change in the disease mix across the country of interest and the reference country. That leads to the following elementary observation:

Observation 1. *Assume that in the country of interest, the disease mix is weighted in favour of diseases with relatively equal death rates across gender. Then*

$$mw_A - mw_B > 0.$$

The opposite inequality would hold if the disease mix is weighted in favour of diseases with higher relative male death rates in the country of interest.

The observation suggests that, if mw_B is close to mw_A, then there is no effect due to the changing composition of disease. We note, however, that the converse implication, that if mw_B falls short of mw_A, then there *is* a compositional effect, has to be regarded more cautiously. The reason is that mw_B is obtained via disease-by-disease aggregation. If some specific disease or disease group is not adequately picked up on the disaggregated data, it will cause mw_B to fall short of mw_A even if there is no compositional effect. Therefore a positive answer—that disease composition does matter—is at best indicative and must be supplanted by other evidence. (We will need to address this issue in Section 5.)

3. MISSING WOMEN BY AGE

We now provide estimates of missing women via the age decomposition described in Section 2.3 and in equation (6). Throughout, we use data from the World Health Organization (WHO) and the UN Population Division. See Appendix A for a description of the data.

Recall that the *overall* sex ratios in China and India are similar, around 1.06, while developed countries and sub-Saharan Africa share similar sex ratios that are in the region of 0.96–0.98. This is why Sen and others have placed India/China into one grouping, while sub-Saharan Africa occupies a different category along with the developed regions. Our accounting suggests that a different classification may be in order.

3.1. *Death rates by age*

Table 2 lists (absolute and relative) death rates of males and females by age group, for India, China, sub-Saharan Africa, and the "developed regions". Predictably, sub-Saharan Africa has very high death rates overall. The high rates are present across all ages and are particularly high at infancy. India's death rates by age group are (also predictably) higher than China's, and much higher than those for developed countries.

However, of greater interest are the *relative* death rates by gender. China and the developed countries have a significantly higher rate of relative mortality for males than do India and sub-Saharan Africa. The discrepancy is particularly noticeable up to middle age.

Figure 1 makes this point by graphing the death rate data (shown in boldface in Table 2) for all four regions. Just as in developed regions, male mortality in China exceeds female mortality at all ages, and the excess is quite pronounced up to the early 1930s. We also see the familiar hump through adolescence and the 1920s, with a particular (relative) surfeit of male deaths.

The contrast with India and sub-Saharan Africa is remarkable. Young ages display an excess of *female deaths*, that excess reaching a peak in near-mirror sychronization to the

TABLE 2

Death rates by age, 2000

	Developed countries			India			China			Sub-Saharan Africa		
Age	Male	Female	Relative	Male	Female	Relative	Male	Female	Relative	Male	Female	Relative
<1	6.1	5.0	**1.21**	76.6	76.2	**1.01**	31.9	37.4	**0.85**	141.5	119.9	**1.18**
1–4	0.3	0.3	**1.25**	5.1	7.7	**0.66**	1.7	2.0	**0.85**	12.7	13.9	**0.91**
5–9	0.2	0.1	**1.31**	1.7	2.5	**0.71**	0.6	0.5	**1.19**	2.8	3.0	**0.93**
10–14	0.2	0.1	**1.49**	1.3	1.5	**0.89**	0.5	0.4	**1.49**	1.9	2.0	**0.95**
15–19	0.7	0.3	**2.37**	2.0	2.4	**0.81**	1.2	0.5	**2.44**	3.7	4.3	**0.86**
20–24	1.1	0.4	**2.94**	2.6	3.3	**0.80**	1.4	0.6	**2.25**	5.7	9.4	**0.61**
25–29	1.1	0.4	**2.60**	3.2	3.2	**1.00**	1.4	0.8	**1.65**	10.8	14.8	**0.74**
30–34	1.2	0.6	**2.20**	3.8	3.1	**1.23**	1.5	1.1	**1.39**	16.4	17.4	**0.95**
35–39	1.7	0.9	**1.92**	5.2	3.3	**1.61**	1.9	1.5	**1.27**	18.7	16.1	**1.16**
40–44	2.5	1.4	**1.82**	6.6	4.6	**1.44**	2.8	2.1	**1.34**	20.2	14.6	**1.38**
45–49	3.9	2.1	**1.86**	9.5	6.2	**1.55**	4.6	3.2	**1.41**	21.7	13.8	**1.57**
50–54	5.7	3.1	**1.84**	13.7	9.6	**1.43**	7.5	5.0	**1.50**	23.8	15.7	**1.52**
55–59	8.9	4.7	**1.90**	22.2	15.1	**1.47**	12.2	7.7	**1.59**	26.9	19.0	**1.42**
60–64	14.0	7.2	**1.95**	31.1	22.2	**1.40**	19.7	12.8	**1.55**	37.2	26.5	**1.41**
65–69	22.6	11.5	**1.96**	47.2	38.4	**1.23**	31.6	21.8	**1.45**	51.6	39.2	**1.32**
70–74	36.0	19.5	**1.85**	68.9	62.1	**1.11**	51.1	39.0	**1.31**	76.7	61.7	**1.24**
75–79	57.2	34.0	**1.68**	99.8	83.5	**1.20**	82.6	69.5	**1.19**	114.6	94.9	**1.21**
80–84	93.7	61.7	**1.52**	126.3	114.0	**1.11**	133.3	118.7	**1.12**	169.8	144.6	**1.17**
85–89	150.1	108.6	**1.38**	166.5	159.1	**1.05**	205.7	191.3	**1.07**	245.7	213.9	**1.15**
90–94	231.9	182.5	**1.27**	228.7	226.6	**1.01**	303.3	291.3	**1.04**	345.3	307.3	**1.12**
95–99	345.8	292.5	**1.18**	327.5	329.7	**0.99**	427.7	418.9	**1.02**	470.3	428.0	**1.10**
100+	501.8	453.9	**1.11**	488.3	489.8	**1.00**	576.5	568.8	**1.01**	616.8	577.2	**1.07**

Notes: "Male", male death rates, "Female", female death rates, both per 1000, "Relative", Male/Female.
Source: World Health Organization and UN Population Division.

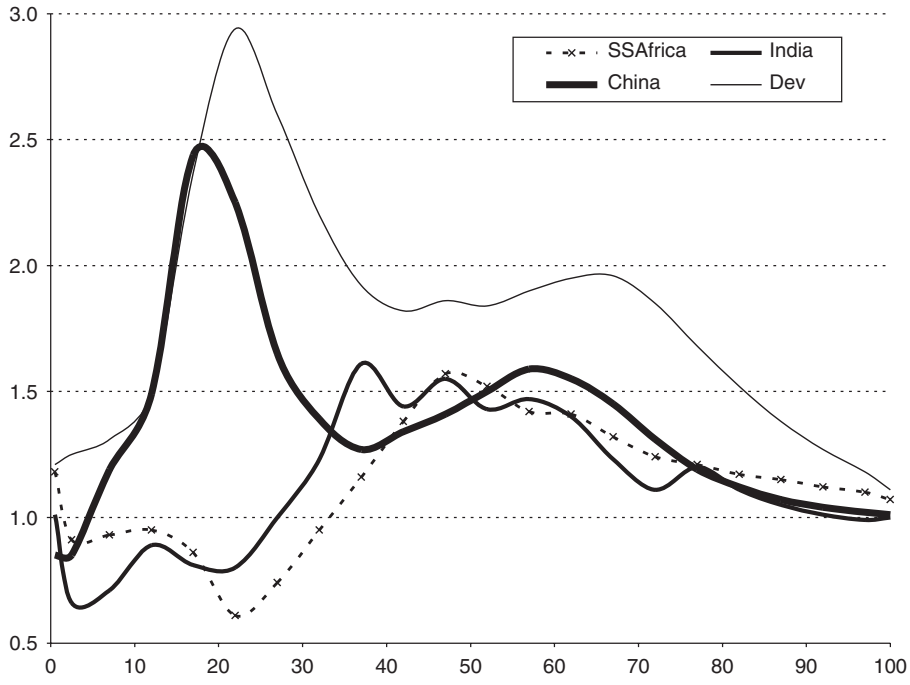

FIGURE 1

Male–female relative death ratios by age

China/developed region pattern. It is only around the mid-30s or later that these relative patterns fall into line, with excess male mortality in all regions. Thus the Chinese profile—barring the ominously higher sex ratio at birth—is not that different from that of a developed country. In contrast, the Indian and sub-Saharan African profiles are similar, at least after birth. The pairings are very different when one looks at age *distributions* rather than overall sex ratios. Other data sources echo these findings.[23]

3.2. *Missing women*

We now form an estimate of missing women by age, using equations (2), (5), and (6). Table 2 has much of the central data we need to accomplish this calculation. We augment these with data on sex ratios at birth, as discussed in Section 2.5.

23. For India, see the National Family Life Survey (2005–2006). For the most recent cohort (born between 2000 and 2006), the relative death rate for ages less than 1 is 1.052; for children aged 1 to 4, it is 0.782. The number of observations is small though: 1500 to 1700 deaths for each sex at ages 0–1 and 300 to 600 at ages 1–4. For sub-Saharan Africa, Garenne (2003) examines child mortality rates using 60 DHS surveys. The relative death rate for ages 0 to 4 is 1.10. The UN Demographic Yearbooks provide information for ten sub-Saharan African countries, and also confirm the relatively low death rates for ages 15 to 29, for the majority of these countries, including Namibia, Bostwana, South Africa, Swaziland, and Zimbabwe. This pattern stands in stark contrast to that for China and the developed countries. "Stock" data on populations reinforce these findings. China and the developed countries have similar profiles: sex ratios decline significantly with age. In contrast, India and sub-Saharan Africa display a relatively flat sex ratio with age. Indeed, India's sex ratios actually *increase* over the age range 0–35, suggesting not just excess female deaths relative to some developed-country trendline, but an absolute excess of female deaths.

For India, recall that sex ratios at birth lie between 1.070 and 1.078, with an average of 1.074 over the years 1996–2006. If we use the developed-country median (1.059) as a reference, the number of missing females at birth in India is approximately 184,000 for the year 2000.[24]

Consider some variations around this estimate. The high-end birth ratio of 1.078 (along with the same reference) gives us 233,000. On the other hand, if we were to use the average of 1.074 and compare it to the actual sex ratio at birth of 1.066 for Indians in the United States, we get around 98,000. Given the mounting evidence for sex selection in the United States, this is surely on the low side. We feel reasonably comfortable with the 184,000 estimate using developed-country reference points.

For China, the estimated sex ratio at birth of 1.169 from the 2000 census and a reference sex ratio of 1.059 together yield 885,000 missing females at birth in 2000.[25] This may be on the high side if the Chinese census systematically under-reports female births. The lowest conceivable number is obtained by using the lower UN estimate for China in 1989, which is 1.139, and employing the implausibly high reference of 1.074, which is the sex ratio at birth for Chinese in the United States, but this still gives us 516,000 missing females. A reasonable compromise may be to use the UN estimates along with the developed-country reference ratio of 1.059, which yields a figure of 644,000. This is probably on the low side, because the UN estimate is for 1989, but we will go with it.

There is substantial variation within sub-Saharan Africa, but Garenne's authoritative review (2002) based on surveys for two-thirds of the region leads to an overall average of 1.033. Our reference range lies between 1.030 and 1.035. Faced with these numbers, it is extremely hard to impute *any* missing females at birth to sub-Saharan Africa as a whole. Perhaps careful disaggregation of the region (along with the corresponding reference ratios) will reveal more nuanced findings in the future, but this is the best that can be done with the data at present.

Table 3 puts these averages together with the estimates of missing females at different ages. Perhaps the most striking implication of the table is that the annual number of missing women in sub-Saharan Africa is comparable to that in India and China, and as a proportion of the total female population it is substantially larger. This is in sharp contrast to much of the earlier literature, which argues that the overall sex ratio in sub-Saharan Africa is low because males and females are treated more equally there.

We are not the first to assert that there are missing women in sub-Saharan Africa. Klasen (1996)[26] points out that the sex ratio at birth is low in sub-Saharan Africa, and uses an adjusted aggregate reference ratio. The numbers, however, are far lower (relative to India and China) than what we obtain here.[27]

We uncover these results precisely because we decompose the problem by age groups. (We will get a clearer picture when we decompose by age and disease.) Note that we impute no missing females at birth to sub-Saharan Africa. The female deficit occurs at later ages.

24. To estimate $\pi^w(0)$, we use the UN birth rate for India (2000) of 25.8 births per 1000. Given the WHO estimate of total population of 1,049,549,480, this implies that there were 27,078,377 births of which 13,018,450 were female.

25. We use the UN estimated birth rate for China in the year 2000 of 14.0 births per 1000 population. Given the WHO estimate of total population of 1,302,307,080, this implies that there were 18,232,289 births of which 8,519,761 were female.

26. See also Klasen and Wink (2003, table 3).

27. However, Klasen and Wink estimate stocks while we estimate flows. As the discussion in Section 6.3 below suggests, the Chinese *stock* of missing women is likely to be much higher than that for India and sub-Saharan Africa.

TABLE 3

Excess female deaths by age (in 000s), 2000

Age group	India	China	ssAfrica
At birth	184	644	0
0–1	146	109	32
At birth + 0–1	330	753	32
1–4	164	23	160
5–9	62	2	40
10–14	31	−0	30
15–19	77	−1	98
20–24	102	7	222
25–29	79	18	258
30–34	50	24	195
35–39	17	26	103
40–44	27	23	47
45–49	24	33	24
50–54	41	28	25
55–59	56	29	35
60–64	86	53	43
65–69	155	100	57
70–74	188	150	62
75–79	112	185	50
80–84	72	151	30
85–89	32	83	11
90–94	9	31	2
95–99	1	6	0
100+	0	1	0
Total (mw_A)	1712	1727	1526
% Female population	0.34	0.31	0.47

Sources: United Nations, WHO, and Table 1. Numbers do not sum to total because of rounding error.

For India and China, we could have chosen the plausibly higher numbers of missing females at birth: 233,000 and 885,000, respectively. This modification increases the flow percentage of missing females (relative to the female population) to 0.35 and 0.31 respectively, still way below the corresponding number (0.47) for sub-Saharan Africa.[28]

Table 3 also reveals that, while India and China are quite similar in the overall numbers and percentages of missing women, their distributions across age in the two countries are quite different. Figure 2 summarizes this observation by plotting the percentages of missing women that can be "attributed" to different age groups in the three regions. China exhibits a huge spike for missing females at birth: 37% of all missing women in China can be attributed to prenatal factors. That number would be as large as 45% if we used the plausibly larger estimate, and no less than 32% even if we use the implausible lower bound.

In contrast, under 11% of the missing females in India are prenatal (at most 13% if we use the higher estimate). The *cumulative* fraction of missing women in India and sub-Saharan Africa does not add up to the Chinese deficiency at birth until the age category of the mid-20s

28. Age-specific mortality rates by gender available for a few sub-Saharan African countries from recent UN Demographic Yearbooks allow us to compute the total number of missing women in those countries. We find that, consistent with the estimates here, the percentage of women missing is in the range of 0.35–0.63 for those countries.

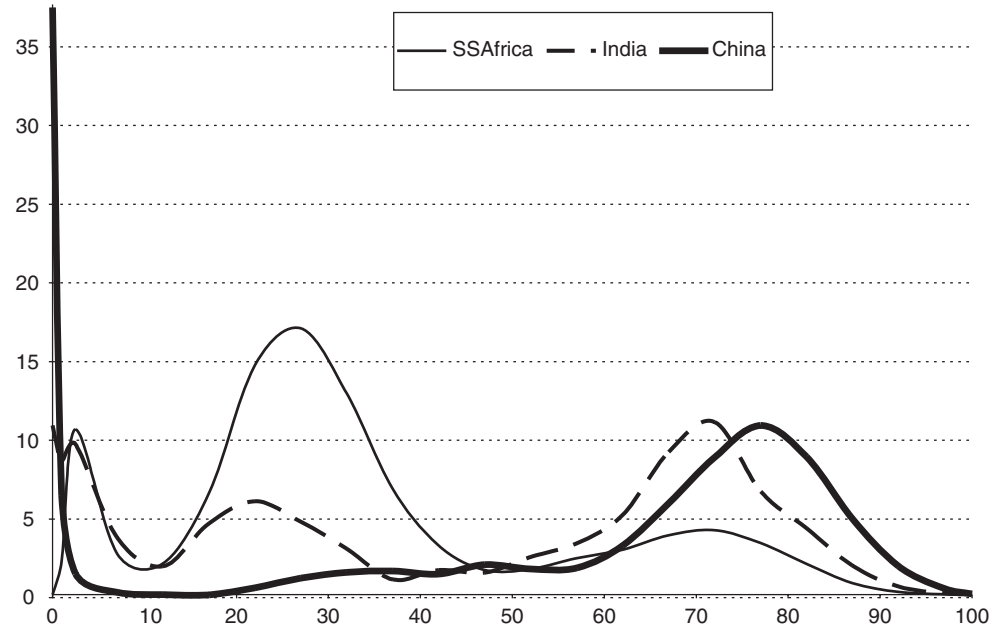

FIGURE 2

Missing women distributed by age (in %)

(or the mid-30s, with the larger Chinese estimates). Fully two-thirds of the missing women in India are from an age older than 15. For sub-Saharan Africa, the proportion of missing women older than 15 is much larger: over 80%.[29]

This is not to suggest that the emphasis on infanticide and sex-selective abortion in India is unwarranted. There are regions in India where such factors play a very important role. The point is that there has been a *relative* neglect of the older age categories, which account for many more missing women. The story is, however, very different for China and fully supports the greater emphasis on prenatal factors, infancy, and early childhood.

3.3. *A remark on pre- versus post-natal mortality*

The distinction between prenatal and infant mortality may be fuzzy in practice. Not all of the missing females at birth can be chalked up to prenatal procedures such as sex-selective abortion. It is entirely possible that female infanticide goes unreported. The missing child would still turn up, but as a missing girl at birth, and not as a post-natal female death. For this reason, an approximate (but useful) measure of missing females "around birth" may be obtained by summing the first two rows in Table 3. We therefore report, in the third row of that table, the total of the first two rows. By this measure, close to 44% of China's missing women are located "around birth", while the corresponding figure for India is under 20%. (The number for sub-Saharan Africa is negligible, around 2%.)

29. The number of missing women in the oldest age groups is significantly smaller in sub-Saharan Africa relative to India and China. This comes from the fact that life expectancy is much lower there and hence the relevant population numbers are correspondingly smaller. The *fraction* of women at older ages who are "missing" is quite similar to those in China and India.

TABLE 4

Excess female deaths in sub-Saharan Africa with two reference groups (in 000s), 2000

Age group	RefGroup = Developed	RefGroup = Afr-Americans
At birth	0	0
0–1	32	45
1–4	160	139
5–9	40	39
10–14	30	31
15–19	98	109
20–24	222	226
25–29	258	254
30–34	195	173
35–39	103	83
40–44	47	32
45–49	24	14
50–54	25	25
55–59	35	34
60–64	43	29
65–69	57	28
70–74	62	33
75–79	50	28
80–84	30	18
85+	9	−3
Total (mw_A)	1521	1337
% Female population	0.47	0.41

Sources: United Nations, WHO, Vital Statistics of the United States, and Table 1. Numbers do not sum to total because of rounding error.

3.4. *A remark on the procedure followed for sub-Saharan Africa*

The figures we report for India and China use a reference sex ratio at birth of 1.059, which corresponds to the developed-country median. In contrast, for sub-Saharan Africa, we use *African* ethnicities residing in developed countries to construct a lower reference sex ratio at birth of 1.033. There is a potential inconsistency here for sub-Saharan Africa: after all, our reference *death* rates do not incorporate variations by ethnicity.[30] Instead, our reference group—as far as deaths are concerned—is a representative sample of all races residing in the different regions of the developed world.

We just do not have mortality data for all of the developed regions disaggregated by race. Even if we did, we could not be sure that gender variations here are due to biology and not discrimination or culture. The process leading to a death is complex, and we lack the controls to understand this. But the process leading to birth is (relatively) simpler, and biological differences stand out better. Of course, it is logically *possible* that African-Americans in the United States have a lower sex ratio at birth because (for instance) they practice sex-selective abortion against males! We would all agree, however, that this argument would be intuitively absurd. Thus allowing for reference differences at birth while shutting them off at death is not necessarily an inconsistent exercise.

Nevertheless, and in the interests of full symmetry, one can use age-specific mortality rates of African-Americans residing in the United States to compute missing women by age in sub-Saharan Africa. Table 4 records the findings. This alternative benchmark reduces the

30. We are grateful to an anonymous referee for raising this point.

total number of missing women in sub-Saharan Africa by 12%, and the percentage of the female population which is missing in sub-Saharan Africa drops by 0.06 points. However, all of our main results still hold. That is, the percentage of the female population missing in sub-Sahara Africa (0.41%) is still much larger than in China or India (0.31 and 0.34%, respectively). The distribution of missing women across the different age groups also remains quite similar.[31]

4. MISSING WOMEN BY DISEASE

We have seen that missing women are present in a variety of age categories in India and sub-Saharan Africa, while such dispersion is far less pronounced for China. We now take a closer look by accounting for missing women over age and disease groups.

The WHO divides the causes of death into three categories: (1) communicable, maternal, perinatal, and nutritional diseases; (2) non-communicable diseases; and (3) injuries. Infectious disease as well as nutritional and reproductive ailments—the Group 1 diseases—predominate in higher mortality populations. These are replaced by chronic and degenerative diseases (Group 2) in low-mortality populations, such as cardiovascular ailments or cancer.[32] This change in mortality patterns, with chronic and degenerative disease replacing acute infectious and deficiency diseases as the leading causes of death, is referred to as the *epidemiological transition* (Omran, 1971).[33] It is easy to see the transition in the data across our three regions of interest; see Supporting Information to this paper for a discussion.

Below, we provide a detailed description of missing women by age and disease category for India, sub-Saharan Africa and China. In light of the epidemiological transition, we also examine the possibility that a divergence in sex ratios between developing and developed countries is simply due to changing compositions of disease. If true, such excess deaths would arise not from lack of "similar care" for men and women but from the changing nature of the disease environment. There is a *prima facie* case for such an argument, simply because Group 1 diseases might discriminate less between males and females, while there may be intrinsic female advantages in surviving Group 2 disease (see Supporting Information for more discussion). This possibility might account for some of the "missing" women.

We use data from the Global Burden of Disease (GBD) study, initiated in 1992, which is a major collaborative effort between the Harvard School of Public Health, the WHO, and the World Bank.[34] The GBD study used numerous data sources and epidemiological models to estimate the first comprehensive worldwide cause-of-death patterns in 14 age–sex groups for over 130 important diseases. The estimates reflect all of the information currently available to the WHO. Appendix A describes this dataset in more detail, and points out some of its limitations.

31. It is important to emphasize that relative to developed regions, 0.08% of the black female population in the United States is also missing. It is this estimate that exactly explains the overall decline in missing women in sub-Saharan Africa with this alternative benchmark. Therefore, to argue that blacks in the United States form a better reference group, we have to be certain that the reasons for "missing" black women in the United States (relative to developed regions) stem entirely from biological differences and not from the possibility that black women in the United States suffer more discrimination than their white counterparts in developed countries. Likely, it is a combination of these two factors (and perhaps others) which explain missing black women in the United States.

32. In contrast, deaths from injury tend to be the most variable across countries, as well as across communities within countries. War deaths represent a typical source of variation.

33. The implicit linearity in this definition is overly simplistic: several infectious diseases, such as tuberculosis, can certainly make a reappearance at "advanced" stages of economic development, along with entirely new infectious diseases, such as AIDS.

34. Refer to Mathers *et al*. (2004) for details.

4.1. *Special methodological considerations*

Recall (3) from Section 2, which yields a "reference" death rate for women by age and disease:

$$u^w(a,k) = \frac{d^m(a,k)}{\widehat{d}^m(a,k)/\widehat{d}^w(a,k)}, \tag{8}$$

and the corresponding expression for excess female deaths by age and disease:

$$\mathrm{mw}(a,k) = \left[d^w(a,k) - u^w(a,k)\right]\pi^w(a),$$

where $\pi^w(a)$ is the starting population of women of age a. If we simply add these numbers up, we obtain an estimate for missing women that *excludes* compositional effects due to the epidemiological transition (refer to Section 2.6 and especially Observation 1):

$$\mathrm{mw}_B = \sum_{a=1}^{n}\sum_{k}\mathrm{mw}(a,k) + \mathrm{mw}(0).$$

There are some exceptions to this approach. The most significant of these is maternal mortality, for which a male death rate is not defined, so equation (8) is invalid. We therefore construct the reference death rate for maternal mortality in each age group by using the ratio of maternal to overall female mortality in each age group in the reference region, and then scaling this by age-specific female mortality for the country in question. That is,

$$u^w(a,\mathrm{mm}) = \frac{\widehat{d}^w(a,\mathrm{mm})}{\widehat{d}^w(a)}d^w(a), \tag{9}$$

where the index $k = \mathrm{mm}$ stands for maternal mortality. Maternal mortality is very low in developed regions, so that this procedure will treat practically all maternal deaths as excess female deaths, which is as it should be.

A second set of exceptions concerns diseases for which relative death rates for developed countries by age are unreliable, because there are so few deaths. Yet those diseases are widespread in the country of interest. Particularly important examples are malaria, childhood cluster diseases (such as measles), diarrhoeal diseases, and tuberculosis.[35] Small changes in the developed-country numbers can cause large swings in our estimates.[36] We therefore use the *overall* death rates from all infectious diseases (excluding HIV/AIDS and other sexually transmitted diseases (STDs)) by age group in developed regions to compute our reference rates.[37]

We have a similar problem for deaths from nutritional causes, HIV/AIDS, and other STDs in the youngest age group, where the death rates from these diseases in the developed regions are close to zero. In these cases, we have nothing to base our estimates on and simply use a reference death ratio of 1:1 as a benchmark.[38]

35. We consider all such categories for which there are at least 2000 female deaths in our country of interest.

36. For malaria, the total number of deaths over all ages and over *all* developed regions was less than 100 in the year 2000. A similar observation is true of childhood cluster diseases. For diarrhoeal diseases and tuberculosis, the situation is somewhat different: there are a substantial number of deaths recorded in developed regions for these two categories of disease, but these primarily occurred at ages 60 or older. Yet in less developed regions, younger age categories account for a large number of deaths from these diseases, particularly in the case of diarrhoeal deaths. We therefore cannot form reliable reference ratios from developed regions in the younger age categories in this case.

37. For a given disease–age category, we consider fewer than 100 female deaths in developed regions to be too small to form reliable reference death ratios.

38. Alternatively, following the same strategy as above, we could have instead used the overall death rates from all communicable diseases within each age group in developed countries to compute our reference death ratios. This exercise does not alter our estimates substantially.

Ideally, medical data could yield reference mortality rates by gender for those diseases for which developed-country data is sparse. However, there is an implicit circularity here: to trust the medical estimates, which are often obtained in developing countries, one must believe, *a priori*, that there is no gender bias in those countries to begin with. We leave these interesting issues as possible topics of future research.

4.2. *General observations*

We report our estimates in Tables 5 (India), 6 (sub-Saharan Africa), and 7 (China). In each table, the main headings are Groups 1 and 2 diseases, and Injuries. The numbers in bold, along the corresponding rows, report missing women by age. In a world with accurate data (including reliable reference death rates) for every conceivable disease, these group sub-aggregates would be built by adding up all missing females from the diseases in that group. Here, they are not. For instance, the Group 1 estimates are found by treating all Group 1 diseases as a *single*

TABLE 5

Excess female deaths by age and disease (in 000s); India, 2000

Disease group | Age	0–4	5–14	15–29	30–44	45–59	60–69	70–79	80+
1. Group 1	**263**	**33**	**61**	**47**	**18**	**37**	**52**	**22**
A. Infectious and parasitic	121	26	16	−4	6	−3	8	6
Tuberculosis	*0**	*0**	*7**	*28*	*17*	*−3*	*0*	—
HIV/AIDS	*0***	*0***	*0*	*−10*	*−1*	—	—	—
Other STDs	*11***	—	—	—	*−1***	*−3***	—	—
Diarrhoeal	*26**	*0***	—	—	*0*	*1*	*0*	*0*
Childhood cluster	*20**	*3**		*2**	—	—	—	—
Meningitis	*6*	*3*	*−1*	—	—	—	—	—
Malaria	*3**	—	—	—	—	—	—	—
Other infectious diseases	*52*	*22*	*2*	*−12*	*3*	*15*	*17*	*7*
B. Respiratory	81	5	0	−2	1	28	37	15
C. Maternal	—	—	65	66	—	—	—	—
D. Perinatal	38	—	—	—	—	—	—	—
E. Nutritional	9**	2**	−1**	0**	14	9	2	0
2. Group 2	**37**	**15**	**44**	**21**	**87**	**178**	**250**	**59**
A. Malignant neoplasms	2	1	4	0	28	21	23	29
B. Diabetes	—	—	—	—	2	8	1	−7
C. Neuropsychiatric	0	2	2	−1	2	1	5	−6
D. Cardiovascular	3	3	19	19	71	160	175	12
E. Respiratory	2	1	4	5	9	2	30	19
F. Digestive	17	8	15	10	16	7	4	−4
G. Congenital	13	—	1	–	—	–	—	—
3. Injuries	**20**	**17**	**86**	**32**	**34**	**22**	**16**	**2**
A. Unintentional	20	15	57	24	24	18	13	3
B. Intentional	0	2	29	8	10	3	2	0
mw_B = **1637**	**320**	**64**	**191**	**100**	**139**	**236**	**318**	**83**
mw_A = **1712**	**310**	**93**	**258**	**93**	**120**	**241**	**300**	**113**

Notes: Figures are rounded to the nearest thousand. "*" implies that the reference death ratios are computed from an average across all infectious diseases in that age group. "**" implies that a reference death ratio equal to 1:1 is used. "–" means that no numbers were reported because female deaths in India totalled less than 2000 in this category. mw_B calculated by adding the numbers for Groups 1, 2, and 3 by age; both mw_A and mw_B also include 184,000 missing women at birth, as in Table 3.
Source: Global Burden of Disease (2002).

TABLE 6

Excess female deaths by age and disease (in 000s); sub-Saharan Africa, 2000

Disease group | Age	0–4	5–14	15–29	30–44	45–59	60–69	70–79	80+
Group 1	**276**	**46**	**402**	**289**	**67**	**−9**	**9**	**2**
A. Infectious and parasitic	270	31	296	221	54	−22	−1	0
Tuberculosis	*0**	*0**	*−9**	*9*	*1*	0	*−1*	—
HIV/AIDS	*−3***	*−1***	*277*	*240*	*78*	*13*	*3*	—
Other STDs	*−5***	—	*11***	*−2***	*−13***	—	—	—
Diarrhoeal	*30**	—	—	—	0	*1*	*−1*	*−2*
Childhood cluster	*54**	*5**	*2**	*1**	*1**	—	—	—
Meningitis	*0*	*2*	—	—	—	—	—	—
Malaria	*138**	*1**	*4**	*5**	*6**	*2**	*1**	*0**
Other infectious disease	*24*	*36*	*21*	*−5*	*0*	*−20*	*−3*	*1*
B. Respiratory	−33	15	31	14	8	3	6	3
C. Maternal	—	—	128	98	15	—	—	—
D. Perinatal	−20	—	—	—	—	—	—	—
E. Nutritional	−2**	1**	—	—	0	0	−2	−2
Group 2	**−3**	**2**	**15**	**0**	**71**	**108**	**112**	**23**
A. Malignant neoplasms	—	0	1	−1	11	11	10	0
B. Diabetes	—	0	—	1	7	10	7	0
C. Neuropsychiatric	0	0	3	0	0	0	−1	−2
D. Cardiovascular	1	2	8	11	55	77	79	22
E. Respiratory	0	—	3	−2	−6	−2	4	3
F. Digestive	—	—	2	0	4	6	1	−1
G. Congenital	−2	—	—	—	—	—	—	—
Injuries	**1**	**2**	**−12**	**−12**	**−4**	**−2**	**−1**	**−0**
$mw_B = 1,385$	**275**	**50**	**406**	**278**	**134**	**97**	**120**	**25**
$mw_A = 1,526$	**192**	**70**	**578**	**345**	**84**	**101**	**112**	**44**

Notes: Figures are rounded to the nearest thousand. "*" implies that the reference death ratios are computed from an average across all infectious diseases in that age group. "**" implies that a reference death ratio equal to 1:1 is used. "−" means that no numbers were reported because female deaths in sub-Saharan Africa totalled less than 2000 in this category. mw_B calculated by adding the numbers for Groups 1, 2 and 3 by age.

Source: Global Burden of Disease (2002).

TABLE 7

Excess female deaths by age and disease (in 000s); China, 2000

Disease group | Age	0–4	5–14	15–29	30–44	45–59	60–69	70–79	80+
Group 1	**129**	**2**	**−3**	**2**	**−7**	**−12**	**−5**	**16**
A. Infectious and parasitic	11	1	−2	−1	−1	−15	−17	−12
Tuberculosis	—	—	*−1*	*7*	*14*	*1*	*−1*	*−1*
HIV/AIDS	—	—	—	—	—	—	—	—
Other STDs	—	—	—	—	—	—	—	—
Diarrhoeal	*8**	—	—	—	—	—	—	*−1*
Childhood cluster	*2**	*0**	—	—	—	—	—	—
Meningitis	*1*	—	—	—	—	—	—	—
Malaria	—	—	—	—	—	—	—	—
Other infectious diseases	*1*	—	—	—	—	—	—	—
B. Respiratory	64	2	−1	−1	−6	−1	7	27
C. Maternal	—	—	4	6	—	—	—	—
D. Perinatal	52	—	—	—	—	—	—	—
E. Nutritional	—	—	—	—	—	—	—	—
Group 2	**17**	**1**	**−1**	**8**	**38**	**111**	**303**	**202**
A. Malignant neoplasms	2	0	−4	−25	−49	−13	26	17
B. Diabetes	—	—	—	1	4	8	10	1
C. Neuropsychiatric	—	—	2	1	1	1	3	7
D. Cardiovascular	—	—	1	9	64	81	153	60
E. Respiratory	—	—	—	2	3	34	123	178
F. Digestive	11	—	−1	0	2	6	6	−1
G. Congenital	5	1	0	—	—	—	—	—
Injuries	**12**	**4**	**14**	**47**	**35**	**12**	**12**	**5**
A. Unintentional	12	3	−4	15	10	2	3	3
B. Intentional	0	1	18	32	24	10	10	5
mw_B = **1592**	**158**	**7**	**10**	**57**	**65**	**111**	**311**	**223**
mw_A = **1727**	**132**	**2**	**24**	**73**	**89**	**154**	**336**	**272**

Notes: Figures are rounded to the nearest thousand. "*" implies that the reference death ratios are computed from an average across all infectious diseases in that age group. "**" implies that a reference death ratio equal to 1:1 is used. "−" means that no numbers were reported because female deaths in China totalled less than 2000 in this category. mw_B calculated by adding the numbers for Groups 1, 2 and 3 by age; both mw_A and mw_B also include 644,000 missing women at birth, as in Table 3.

Source: Global Burden of Disease (2002).

ailment, and then calculating all missing females for that composite ailment. The same is true of Group 2 diseases and Injuries.[39]

However, the total, reported as mw_B on the penultimate line of each table, is deliberately obtained by simply summing over these three categories. If we did not do this, and treated *all* diseases as one composite ailment, we would be entirely unable to separate out the influence of a change in the disease composition. By forcing ourselves to add over these categories, we freeze the disease composition to that in the country of interest, and therefore pick up the "within-disease" component of missing women. The epidemiological transition is located in the difference $mw_A - mw_B$, as Observation 1 makes clear.

4.3. *India*

Table 5 summarizes our results for India. The very last line recalls our earlier estimates for missing women by age (mw_A). The penultimate line records mw_B, as discussed above, providing estimates for every age category as well as a total, obtained by adding in the estimate for missing girls at birth. The correspondence between the two rows is remarkable. Recall that mw_B deliberately eliminates the effect of the changing composition of disease across developed and developing countries, while mw_A includes all changes in disease composition. *Yet there is little difference between the two sets of totals at most ages, and the two grand totals mw_A and mw_B practically agree.* By Observation 1, we must conclude that few, if any, of the missing women in India can be attributed to the epidemiological transition.

Several additional observations are of interest. First, it is evident that the bulk of missing females at younger ages come from Group 1 diseases. Group 1 disease accounts for fully 260,000 missing females between the ages of 0 and 4, which is over 15% of the total. Of these, about half is due to infectious disease, while the remainder may be attributed to respiratory and perinatal ailments. To provide some idea of how big this number is, consider maternal mortality in India, which is widely acknowledged to be a serious issue (see, e.g. Ronsmans and Graham, 2006). Maternal deaths account for about 130,000 excess female deaths, no small number, but of the same magnitude as *excess* female deaths caused by infectious and parasitic diseases within the age 0–4 category *alone*.

As we have already seen, much of the Indian discrepancy is to be found at older ages. That excess is certainly reflected in Table 5. At these ages, the excess burden falls mainly on non-communicable Group 2 diseases. Cardiovascular disease is particularly implicated. Women die at a rate closer to men from cardiovascular disease relative to developed countries.[40] Our estimates suggest that the number of excess female deaths from cardiovascular sources in the age category 60–69 alone significantly exceeds all the excess female deaths due to maternal mortality. Moreover, the same number is repeated in the 70–79 age category.

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; Kochar, 1999). Table 5 is fully supportive of this emphasis.

One rather sinister observation is that the number of excess female deaths from "Injuries" is high in India. There are excess female deaths under this heading in *all* groups. Excess female deaths for women from "Injuries" exceed 225,000, a number that dwarfs maternal

39. Still, it should be noted that despite the incompleteness of the data, the numbers obtained by summing across the diseases in each sub-category are, in fact, *not* that far from the group aggregates that we do use. For instance, add items *A* through *E* in Group 1 for India and compare it to the Group 1 aggregates for every age group.

40. For example, the death rates from cardiovascular disease for the age group 70–79 for males and females in India are 26.17 and 22.20, respectively; the corresponding numbers in developed countries are 15.43 and 9.29.

mortality. The category 15–29 stands out in this regard, where the number of excess female deaths from "Injuries" outpaces excess deaths from maternal mortality at the same age. Further decomposition of "Injuries" into "unintentional" (accidents, etc.) and "intentional" (resulting essentially from acts of violence) tells us that around 30,000 *extra* women die per year, of "intentional injuries", or *reported* violence, in the 15–29 age category alone. These are large and disturbing numbers.

Two other factors that point at violence, this time possibly at female infanticide, are the large number of excess deaths under perinatal and congenital conditions. The former accounts for 38,000 excess female deaths and the latter for over 13,000; again, these are large numbers, comparable to excess deaths from "intentional injuries" in the 15–29 age category.

4.4. *Sub-Saharan Africa*

Table 6 reports estimates for sub-Saharan Africa. As in the case of India, we begin with a comparison of the penultimate row, which adds up excess female mortality over the three groups of disease, with the very last row, which records overall excess deaths by age with no thought given to particular disease groups. The difference, as before, proxies the effect of a change in disease composition across the three main groups. Once again, the correspondence between the two sets of numbers is quite strong, and in particular there is no evidence that the epidemiological transition *per se* accounts for too many missing women. Overall, the transition appears to account for under 10% of the total.

The majority of excess female deaths in sub-Saharan Africa fall into the age groups 0–4, 15–29, and 30–44, with a particularly large number in 15–29. The number of missing girls in the youngest age group (equal to 275,000) is comparable to those missing in India of that age (320,000). As in India, diarrhoeal and vaccine-preventible childhood cluster diseases account for a significant number of missing girls in sub-Saharan Africa, but unlike in other regions, malaria is the principal cause of excess female mortality in the age group 0–4. Each year close to 140,000 young girls are missing from this disease alone.

However, by far the overwhelming single cause of excess female mortality is HIV/AIDS. It accounts for over 600,000 excess female deaths, largely in the 15–44 age category. This category is also of special interest because it raises philosophical questions about the benchmark used to measure missing women. Are these 600,000 excess deaths due to "discrimination", violence, or different cultural norms regarding sexuality? For instance, the heterosexual transmission of HIV/AIDS in sub-Saharan Africa (in contrast to several developed countries) and the resulting discrepancy in relative death rates across gender surely account for these excess female deaths. But what lies behind this difference? Section 6 visits this question of interpretation, but the fact is that we do not know the exact channel. What we do draws attention to it in a way that permits comparison with other age-disease groups. HIV/AIDS is a category—one that accounts for an enormous number of deaths—that differentially affects women in sub-Saharan Africa. Ours is a first estimate of just how large that differential might be, relative to other categories.

At the same time, it is important not to underestimate excess female mortality from other causes. If we remove the HIV/AIDS numbers entirely from the sub-Saharan African total, that still leaves us with over 900,000 missing women, which is 0.28% of the female population. This is entirely comparable with the overall percentages for India (0.31) and China (0.34).

As in the case of India, we see that at younger ages, Group 1 diseases play a major role in accounting for missing females, while at older ages Group 2 diseases predominate. The cardiovascular disadvantage for women makes its presence felt, just as it does in India, though the relative contribution of maternal mortality—and certainly HIV/AIDS—is significantly higher for sub-Saharan Africa.

It is important to reiterate that the important causes of death are not necessarily the important repositiories of *excess female* death. Respiratory, perinatal, congenital, and nutritional deaths are important in sub-Saharan Africa, but we do not see an obvious female bias in these deaths. Recall that India, in contrast, displayed a suspiciously large number for excess female mortality under the perinatal and congenital headings, as well as under "Injuries".

4.5. *China*

Table 7 reports analogous estimates for China. As we have seen before, China is different from both sub-Saharan Africa and India. It has a similar number of missing women, but the bulk of them—around 37% and plausibly more, up to 45%—are to be found at birth. Thereafter, the highest numbers occur for the lowest age group (0–4) and then for the three oldest age groups (60 and older). Maternal mortality is not a serious issue. Excess female child mortality is due mainly to Group 1 diseases; particularly those classified under "respiratory" and "perinatal". The disquiet raised by these numbers is not unlike that felt when examining the Indian case under the "perinatal" or "congenital" headings.

For the older ages, Group 2 diseases explain the excess female deaths. The main causes are cardiovascular and respiratory diseases. Once again, the *excess* female deaths occur in China because women die at a rate closer to that of men from these diseases relative to developed countries. Overall, the percentage of missing women due to non-communicable diseases is similar for India and China, though the composition by disease is distinct. In India, there are excess female deaths due to cancer, and in China there are far more women dying from respiratory diseases. Maternal conditions play a role in explaining the number of missing women in both India and sub-Saharan Africa but not in China.

Finally, as far as "Injuries" are concerned, the situation seems to lie further away from sub-Saharan Africa and closer to India, where "Injuries" form a large component of excess female deaths. There are certainly excess female deaths in China under this category. As in India, there are also missing women from "intentional" injuries caused by deliberate acts of violence: the 30–44 and 45–59 categories appear to be particularly hard hit.

As in the case of India and sub-Saharan Africa, disease composition seems to have little to do with excess female mortality in China. In the two major age groups with (post-natal) missing females, there appears to be little or no composition effect. In the age category 0–4, the effect is, if anything, reversed: the changing composition of disease is associated with higher excess female mortality. There is a definite effect in the 60–69 age category, where disease-by-disease comparisons account for a little over two-thirds of the missing women. Otherwise, the epidemiological transition does not appear to account for many missing females. Certainly, if we go by the overall numbers, the transition explains under 8% of all missing women in China.

5. A HISTORICAL EXCURSION: THE UNITED STATES IN 1900

Barring the limitations of data, there is no reason why the approach in this paper cannot be applied to the historical experience of now-developed countries. In this section, we show how this can be done for the United States in 1900.

The profile of age-specific relative death rates supports the contention that males and females died far more equally in 1900 in the United States than they do now. By the same criterion applied to developing countries today, women were at a relative disadvantage then. Figure 3 illustrates. The upper curve plots the male-to-female relative death rates in the developed regions today; it is exactly the same curve as in Figure 1. The lower curve depicts

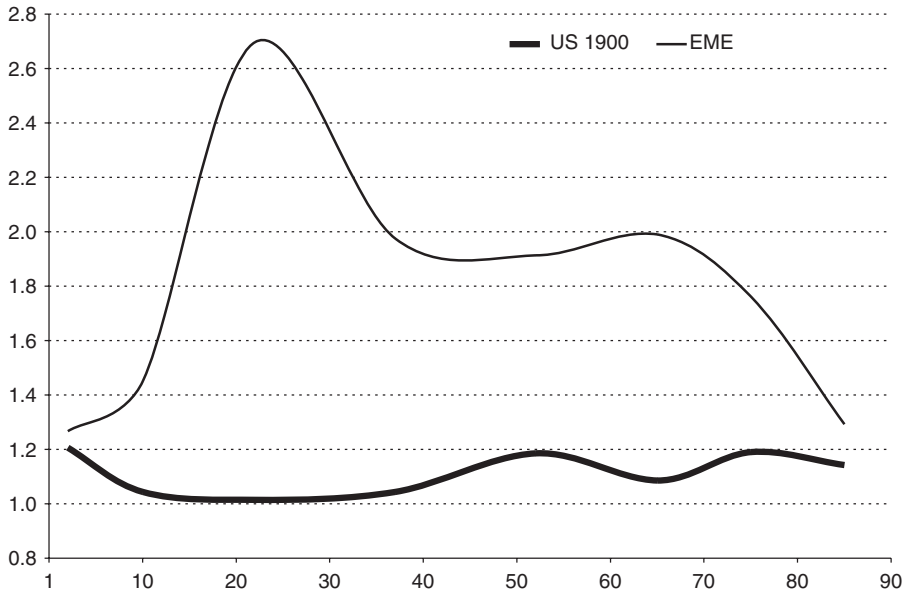

FIGURE 3

Male–female death ratios by age for developed regions and the United States in 1900

Source: UN, WHO, Historical Vital Statistics of the United States.

the same profile for the United States in 1900. As far as relative death rates by age in the United States in 1900 go, the latter then looked much like sub-Saharan Africa and India look today.

The data that go into Figure 3 allow us to quickly form estimates of missing women by age in the United States in 1900. We do so in Table 8, which recalls the three developing regions as well for easy comparison. The table is best read by mentally scaling up the US column by a factor of 10, which puts the relative numbers on par with sub-Saharan Africa. This is because the female population in 1900 in the United States was around 37 million, while the corresponding number for sub-Saharan Africa is around 350 million.[41] Note, too, that the sex ratios at birth for different race/ethnic groups in the United States have not significantly altered between 1900 and 2000, therefore we record zero missing women at birth.[42]

Nevertheless, substantial numbers of women were indeed missing in 1900 in the United States. As a proportion of the female population, the total for 1900 in the United States is actually larger than in India or China today. What is more, with the exception of the youngest age group (0–4), the pattern of missing women in the United States in 1900 is quite similar to that of India and sub-Saharan Africa today (examine the table with the mental scaling of 10).

We now estimate missing women by disease. We rely on the Historical Census Reports from the Vital Statistics department of the United States. There are some differences between the way this data is presented and the counterpart tables for now-developing regions (Tables 5– 7). We have attempted to adhere to the GBD classification used in the earlier tables, but the classification of disease in 1900 was different. Some of this is for obvious reasons: the HIV/AIDS virus was completely absent and others like tuberculosis and meningitis were significantly more

41. A scaling of roughly 15 would put the numbers on par with India, and of around 20 with China.
42. The overall sex ratio at birth in the United States in 1900 is reported to be 1.048; in the year 2000 it is 1.049.

TABLE 8

Excess female deaths in developing regions (in 000s), 2000 and in the United States, 1900

Age group	India	China	ssAfrica	1900 US
At birth	184	644	0	0
0–4	310	132	192	7
5–14	93	2	70	8
15–29	258	24	578	45
30–44	100	73	345	30
45–59	120	89	84	22
60–69	241	154	101	23
70–79	300	336	112	16
80+	113	272	44	4
Total (mw_A)	1712	1727	1526	155
% Female population	0.34	0.31	0.50	0.42

Note: 1900 US female population approximately 37 million.
Sources: United Nations and WHO, Table 1 and US Historical Vital Statistics.

present. But there are other issues of classification.[43] Therefore, it is important to bear in mind that Table 9, which summarizes our results, can only provide rough estimates of the number of missing women by disease.

Nevertheless, the table unearths a pattern of missing women in the historical data similar to that in developing countries today. A notable exception is that there appears to be significantly fewer missing girls (aged 0 to 4) in the historical data. One major reason for this (at least relative to India and China) is that there is very little respiratory female deficit in the historical United States in 1900, while these two deficits are significantly large in India and China. These deficits, by the way, point directly to pre- and post-natal gender discrimination in India and China in a way that does not seem to have been present in the historical United States, and possibly not in modern sub-Saharan Africa.[44]

The table is conservative in its estimates of young missing females due to tuberculosis, which was a major killer in the 1900 United States. Modern developed regions do not have a large enough number of tuberculosis-related deaths at early ages to allow us to form reference death ratios with any degree of confidence. As soon as those reference ratios become reliable (post age 30), we do see a large number of excess female deaths due to tuberculosis. As in Section 4.1, for the younger groups we use as reference the average death ratio from all infectious diseases in developed countries in the relevant age group. If, instead, we use the

43. Among non-communicable disease, apoplexy and Bright's disease were recorded as leading killers. However, apoplexy was used to describe any death that began with a sudden loss of consciousness, especially if death followed soon after. So, for instance, death from cardiac fibrillation, a ruptured aneurysm, and perhaps even some perinatal or respiratory conditions were likely all clumped together. We have included apoplexy under "cardiovascular diseases" in line with the present classification system, but be aware that this probably accounts for at least some deaths in other categories. Similarly, Bright's disease is an older classification for different forms of kidney disease. The term is no longer employed, as the relevant complex of kidney diseases would now be classified by their better understood aetiologies. The same can be said for deaths from "convulsions", which has been placed in the category of "neuropsychiatric conditions" under the present classification.

44. Another difference is reportedly high death rates from non-communicable diseases for ages 0 to 4 in 1900 United States; and there is excess male mortality in this case. This is quite likely a misclassification problem. The two most significant killers in this category are "convulsions" (classified as a neuropsychiatric condition) and "debility and atrophy" (classified as congenital anomalies). Neither of these conditions is listed according to these terms in today's classification.

TABLE 9

Excess female deaths by age and disease (in 00s); United States, 1900

Disease group | Age	0–4	5–14	15–29	30–44	45–59	60–69	70–79	80+
Group 1	**85**	**26**	**168**	**201**	**80**	**68**	**64**	**21**
A. Infectious and parasitic	51	25	176	149	55	19	15	4
Tuberculosis	*3**	*13**	*104**	*124*	*50*	*17*	*11*	*3*
Diarrhoeal	*15**	*8**	*−11**	*17**	*5*	*9*	*3*	*−1*
Childhood cluster	*28**	*9**	*4**	*5**	*2**	*0**	—	—
Meningitis	*4*	*0*	*−2*	*0*	*0*	—	—	—
Malaria	*4**	*2**	*3**	*4**	*2**	*1**	*1**	—
Other infectious diseases	*2*	*1*	*10*	*8*	*0*	*0*	*−1*	*1*
B. Respiratory	10	3	1	13	20	44	45	17
C. Maternal	—	—	54	40	—	—	—	—
D. Perinatal	−4	—	—	—	—	—	—	—
E. Nutritional	−10**	—	2**	1**	2	1	—	—
Group 2	**−21**	**5**	**71**	**98**	**116**	**100**	**62**	**−9**
A. Malignant neoplasms	—	—	3	19	39	24	14	4
B. Diabetes	—	—	0	0	2	1	−1	—
C. Neuropsychiatric	−6	2	12	9	8	10	6	−6
D. Cardiovascular	−3	7	23	41	62	51	28	−8
E. Respiratory	0	—	2	2	−1	2	3	2
F. Digestive	10	−4	15	22	15	10	4	−1
G. Genitourinary	0	1	13	17	−1	−10	−16	−12
H. Congenital	−5	—	—	1	1	1	1	2
Injuries	**6**	**−5**	**−5**	**−8**	**−9**	**−2**	**1**	**3**
mw_B = **1,115**	**69**	**26**	**234**	**291**	**187**	**166**	**127**	**15**
mw_A = **1,548**	**71**	**79**	**454**	**298**	**215**	**233**	**157**	**4**

Notes: Figures are rounded to the nearest hundred. Group 1 and Group 2 totals are compiled from sub-categories (A, B, C, D, not finer sub-categories) and these totals are added to "Injuries" to calculate mw_B at every age group. There are no missing women at birth. "*" implies that the reference death ratio is computed from an average across all infectious diseases in that age group. "**" implies that a reference death ratio equal to 1:1 is used. "–" indicates no numbers reported because female deaths in 1900 United States totalled less than 200 in this category.

Source: Historical Census Reports, National Vital Statistics of the United States.

reference death ratio at ages 30 to 44 (for which we have sufficient data in the developed regions), then these estimates of missing women from tuberculosis at the younger ages in 1900 United States more than double. Relative to that benchmark, it is possible to trace over 26,500 excess female deaths for women aged less than 30, suggesting that there is a significant number of females missing on this score in the historical United States.

Now we study the aggregates. Our historical dataset for the United States does not provide totals for Groups 1 and 2, and we must construct these ourselves as best we can from the sub-categories. The aggregate excess deaths for the two groups are obtained by adding excess female deaths from sub-categories A–D (we do not aggregate finer subdivisions, as data at those levels are just not comprehensive[45]). We then obtain estimates for missing women by age by adding over Groups 1, 2, and "Injuries"; recall that these yield estimates that have been purged of compositional effects, and add up (over age) to mw_B. Table 9 records these, as also our earlier estimates (mw_A) with the compositional effects included.

Unlike in developing regions, there appears to be a significant disease-composition effect. The shortfall, $\text{mw}_A - \text{mw}_B$, is close to 30% of the total, something that was decidedly not the case for the developing countries studied earlier. However, the remarks in Section 2.6 following Observation 1 warn us that this is not a definitive test for a compositional effect. The reason is that the disaggregated data may be missing some disease groups. This is unlikely (because the groups A–D are large, inclusive sub-groups), but possible. The Supporting Information to this paper provides supplementary evidence for a strong compositional effect.

6. MISSING WOMEN? SOME CONCEPTUAL ISSUES

6.1. *The counterfactual*

As in Sen's counterfactual, the implicit presumption is that there are two differences between our country of interest and the reference country. One is "scale": our country may be poorer and so have higher death rates for both men and women. The other is gender bias, reflected in different *relative* death rates. (2) implies that the former effect is not included in our computation of missing women; it is only the latter.

Like all choices of reference standards, ours can be freely criticized.[46] One might argue, for instance, that the *absolute* differences in death rates "should" form the relevant standard. Obviously, low death rates under any category must translate into low absolute male–female differences, and we are generally forced to the conclusion that there is little or no gender differential for reference death rates in that category. For instance, suppose that female and male death rates for some category are 1 per 1000 and 2 per 1000, respectively, in the reference country, while they are 10 per 1000 and 12 per 1000, respectively, in the country of interest. By the "relative" reference standard we use (as do Sen and Coale and others), there are missing women in the country of interest. By the "absolute" standard, there are missing men![47] While it

45. The conceptual advantage of building mw_B from the finest sub-groupings available is, of course, that we purge the totals of any change in disease composition, thereby allowing a cleaner comparison with mw_A. But the disadvantage is, of course, that death information is simply not available over a full partition of the finest categories of diseases, so that there will be many omissions.

46. The comments that follow owe much to helpful discussions with Stefan Klasen.

47. As a related but distinct consideration, we note that (2) embodies another implicit assumption: that the death rate for men in the country of interest is somehow the "right" death rate, and that the female death rate is distorted (upwards). It is possible that *both* death rates are distorted: female upwards, male downwards. In that case, the reference female death rate would be higher, and this would scale down the number of missing women. We are grateful to Jean-Marie Baland for making this point.

is important to keep these alternatives in mind, we prefer the scale neutrality that is embodied in the use of relative death rates.

Such a discussion about the appropriate functional specification is related to a broader issue. How do we know that gender-based death rates are not somehow "naturally" different at different levels of development? The epidemiological transition suggests that they may well be, once aggregated over disease. But we have already seen that compositional effects are minimal: missing women aggregated disease by disease (with compositional changes removed) account for almost all of the grand total.

That still leaves us with the rebuttal that relative death rates for each age *and* disease category might "naturally" vary with development. But this statement is impossible to test with the data we have. We would have to *presume*—at least for a sizable set of countries once poor, or poor countries today—that there is no gender discrimination.

The reader may nevertheless insist on computations for "missing women" that rely on reference standards from other countries that are poor today, or from now-rich countries when they were poor. For instance, one can recompute missing women by age (extending Table 3) by using Latin American and Caribbean countries as the reference group. The results are given in Appendix B. One might also use life tables based on European data from the 19th and early 20th centuries, but it is important to appreciate that life tables provide parametric estimates based on *actual* death rates prevailing in countries at different levels of development. There should be no presumption that these are somehow "neutral" towards gender in any aspect.

In our opinion, any reference group that does not replicate what we see in developed countries today runs the risk of burying important gender differentials under the cover of a "benchmark". It is conceptually more satisfactory to presume that the "natural" relative death rates are indeed constant with development, and then to view every departure from that benchmark as *prima facie* cause for suspicion (though not as conclusive evidence).

Indeed, we are extremely cautious—and in Section 6.2 below, will be very cautious again—in interpreting our results as providing direct evidence of discrimination. We are open to considering a collection of possible factors that might create the discrepancies we do observe. As the next subsection will make abundantly clear, we do *not* assert that all such departures represent overt (or even implicit) discrimination against women. In general, there will be an entire complex of social, behavioural, and economic pathways that will need to be invoked. Our research agenda is to highlight each of the sources of excess female mortality in a unified and comparable way, so that the relevant pathways can be explored in future research.

But we nevertheless want to retain the ability to fully examine any observed departure from the gender-based death ratios we see *in developed countries today*. We are reluctant to conceal those departures by considering other benchmarks.

6.2. *The pathways*

Estimates of missing women were originally meant to represent some measure of the degree of gender discrimination. That may or may not be the case. Much more work is needed to identify the underlying mechanisms.

There are certainly situations, such as excess female deaths from "Injuries", that appear to serve as definitive indicators of overt violence against women. In South Asia, fire-related death is a leading cause; each year over 100,000 women are killed by fires. In East Asia, self-inflicted intentional injuries is the primary cause of death: well over 100,000 women die

each year from suicide. In India, excess deaths from injuries could well be associated with dowry.[48] Our calculations imply very large numbers for India.

The Chinese case is also suggestive. According to the WHO, China is the only country in the world where women are more likely to commit suicide than men. Women are 25% more likely to commit suicide than men (in rural areas they are 66% more likely). This is surely reflected in our estimates on "Injuries". Most theories focus on the low status of rural women to explain these startling differences (Phillips *et al*., 2002).

Next, consider maternal mortality. A crucial accomplishment of the twentieth century in the developed world is the virtual elimination of maternal mortality. By contrast, as is clear from our estimates, maternal death rates remain high in many parts of the developing world.[49] Whether a woman dies while giving birth depends largely on timely and competent obstetric care. Despite this, a large proportion of maternal deaths in the developing world occur in hospitals (Ronsmans and Graham, 2006).

Is this lack of "similar care"? At one level, the answer is no, because men do not give birth, so the question of similar care does not arise. Yet, in a broader context one might call it just that: the prospect of maternal death exists, while no similar prospect exists for males. "Similar care" might dictate that we eliminate this asymmetry. While we personally subscribe to this view, the situation is somewhat more ambiguous than it is for "Injuries".

Now consider a more involved example. Consider missing women in sub-Saharan Africa as a result of the HIV/AIDS epidemic. The numbers are nothing short of dramatic: as Table 6 demonstrates, there are close to 600,000 excess female deaths each year from the virus (over a third of all missing women in sub-Saharan Africa). These extra deaths mainly occur at ages 15–44. The death rate from the HIV/AIDS virus for women of age 15–29 is 2.3 times that of males of the corresponding age. The overall female death rate from the virus is 1.2 times that for males. Elsewhere in the world, in contrast, the death rate from the virus for males is higher at all ages. The ratio is as high as 4:1 in high-income countries.[50]

To understand the extent to which these extra female deaths correspond to discrimination, it is important to understand the significant male slant in the reference rate. In developed countries, it might be thought that men who have sex with men (MSM) engage in risky, unprotected sex, thereby increasing the chances of transmission. But evidence from the United States indicates that the incidence of unprotected sex is fully comparable for heterosexual and homosexual encounters.[51] MSM intrinsically exhibit higher transmission rates.[52] As the calibrations of Goodreau and Golden (2007) for the United States reveal, "In order for US heterosexual individuals to have a generalized epidemic, they would need to adopt levels of unprotected sex several times higher than those currently exhibited by MSM".

48. Indeed, the custom of dowry has been linked to bride-burning and dowry-death (and violence more generally) if promised dowry payments are not forthcoming (see, for example, Bloch and Rao, 2002). The National Crime Bureau of the Government of India reports approximately 6000 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.

49. Using our historical example, the death rate from maternal causes in the United States in 1900, for women aged 15 to 29, was 0.50 per 1000 women. In South Asia today the rate is similar at 0.53 and in sub-Saharan Africa it is much higher at 1.34.

50. The corresponding ratio is 3 in South Asia and more than 2 in East Asia, the Pacific, Latin America, the Caribbean, the Middle East, and North Africa.

51. See Catania *et al*. (2001) and Goodreau and Golden (2007).

52. Goodreau and Golden (2007) discuss this issue in some detail. Briefly, there are two reasons for the higher transmission rate: penile–anal sex is more susceptible to transmission than penile–vaginal sex, and transmission is higher for receptive partners, and among MSM the same individual often plays both an insertive and receptive role.

In the light of these observations, the high levels of heterosexual transmission in sub-Saharan Africa acquire enormous salience. There is little doubt that poverty, undernutrition, and poor health care play their part in the overall transmission of HIV. But our interest lies not in the overall rate of transmission but in its reversal of male–female incidence. It is, therefore, hard to escape arguments such as those made by the WHO, which emphasize the role of unequal power and violence.[53] It seems evident that the multiplicity of female sexual partners among males, the prevalence of transactional sex, the existence of violent or forced sex, and the relative inability of women to negotiate safer sex practices have much to do with this extraordinary discrepancy.[54]

There *could* be other reasons for greater heterosexual transmission that have nothing explicitly to do with female discrimination. For example, it is possible that certain aspects of poverty can create "unintended" gender biases relative to the developed-country benchmark.[55] The accounting methodology that we follow is entirely silent on matters of interpretation. The case for (or against) lack of similar care has to be made separately.

The cardiovascular deficit presents another sort of interpretative quandary. A key finding is that heart disease accounts for a large fraction of excess female mortality. In developing countries, women simply die of cardiovascular disease at a rate closer to that of men.[56] 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 (Dhandapany *et al*., 2009) may be equally present in males and females, thereby lowering the gender skew in incidence.[57] 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 "similar care": women seek or receive medical care less often in developing countries, or may be subject to greater stress. Understanding why women appear to be at a particular disadvantage from cardiovascular disease is important work for future research.

It is beyond the scope of this paper to disentangle the role of direct gender discrimination from other factors—biological, social, environmental, behavioural, or economic—in explaining the pervasive phenomenon of excess female mortality.[58] This is an important direction for future research and is already a central item on the policy agenda for international development agencies. By moving away from (while not abandoning) recent literature that highlights the importance of the sex ratio at birth as a key determinant of missing women, we have taken a preliminary step towards a unified study of a much broader set of issues. In other words, while some of these sources of female deficits, such as cardiovascular disease, may be known, our methodology puts them into one unified framework where the numbers from sources as diverse as Injuries, HIV/AIDS, cardiovascular, or birth can be put together and compared.

53. For a summary of this position, see http://www.who.int/gender/hiv_aids/en/.

54. On these and related matters, see, e.g. Bongaarts (1996); Maman *et al*. (2000); Gupta (2002); Dunkle *et al*. (2004), and Silverman *et al*. (2007).

55. For instance, poverty and the lack of health care precipitate a higher rate of genital ulceration, which may serve to reduce the gap between MSM and heterosexual transmission rates. Schistosomiasis—infection by a parasitic worm—affects millions in sub-Saharan Africa. In women, it damages the lining of the vagina, which protects against HIV transmission. On schistosomiasis and HIV transmission, see e.g. Chenine *et al*. (2008), though exact magnitudes are harder to quantify because of the bi-directional nature of the relationship (Secor, Karanka and Colley, 2004).

56. See Nikiforov and Mamaev (1998) for a historical perspective on gender imbalances in cardiovascular disease in now-developed countries.

57. The death rates from cardiovascular disease in India are well known to be high. Even in rural areas of India, chronic diseases (primarily coronary) are now the leading cause of death (see Joshi *et al*., 2006).

58. There is a large literature aimed at explaining differential mortality patterns by gender in *developed* countries; see, e.g. Nathanson (1984) for a survey.

6.3. *Stocks and flows*

Unlike several of the central papers we have cited, such as Sen (1990) and Coale (1991), our paper discusses annual "flows" of missing women, rather than the existing "stock" at any point of time. To understand the issues involved in converting flows to stocks, consider the difference between the actual female death rate and the corresponding reference rate:

$$\Delta(a) \equiv d^w(a) - u^w(a).$$

In computing the flow of missing women at age a, we multiply $\Delta(a)$ by the female population at age a, as we have already done. However, the impact of $\Delta(a)$ on the overall stock of missing women is more far-reaching: it affects *every* cohort of women currently older than a. More precisely, a cohort of age a' is diminished at an earlier age a by an amount that depends on the value of $\Delta(a)$ that prevailed $a' - a$ years ago. In general, we lack data on this value and this creates a serious difficulty in converting the flow numbers to stocks. We could make the "steady state" assumption that all age-specific death rates have remained constant over several years, but it would have to be defended, and that is beyond the scope of this paper.

However, it is worth noting that, no matter how we resolve this issue, a given proportional flow at an earlier age *must* be more important than the same flow at a later age. It will simply affect more women, and must therefore translate into larger stocks. Simply as an example, we report our tentative estimates under the assumption of time-invariance of Δ. We find approximately 20 million missing women in India, 58 million missing women in China, and 8 million missing women in sub-Saharan Africa. Look at the enormous difference between China and the other two regions (in flows all three regions were about the same). This comes from the fact that excess female deaths in China are clustered at age 0. We reiterate, though, that these estimates must be treated with a great deal of caution.

7. SUMMARY AND DISCUSSION

Defining missing women by differences in aggregate sex ratios can be misleading, or uninformative (or both). It is misleading because different countries have different fertility and death rates, and (in particular) different age distributions. They will have different disease compositions. They may also have different sex ratios at birth for genetic or environmental reasons that have nothing to do with missing females.

The procedure is also uninformative: we cannot tell at what ages the missing women are clustered, or what diseases are responsible. Thus, we cannot begin to ask about the various channels: discrimination, biology, social norms, and so on. Answering these questions is of profound importance. By unpacking missing women by age and disease, our paper takes a limited and preliminary step in this direction.

Our study of excess female deaths by age yields two key findings. First, once we control for natural variations in the sex ratio at birth, sub-Saharan Africa has as many missing women as India and China: significantly more as a percentage of the female population. Second, the majority of missing women are of adult age. Sub-Saharan Africa has no missing females at birth, while the corresponding proportion for India is under 11%. China's missing females, in contrast, are largely prenatal. About 37–45% of them may be classified as missing at birth. But all these regions display missing women at a variety of ages. For instance, excess female mortality up to age 15 does not account for more than a third of the total in India or sub-Saharan Africa.

Our study of excess female deaths by age and disease yields the following findings. For developing countries today, the epidemiological transition—the changing *composition* of disease—explains very little of excess female mortality. At young ages, the Group 1 diseases are

largely responsible for missing women. The opposite is true at older ages; Group 2 diseases are responsible.

Aside from these commonalities, each region tells a different story. In India, communicable, preventable diseases explain missing girls in childhood. Maternal mortality and injuries are important at the reproductive ages. Cardiovascular deaths are an overwhelmingly strong source of missing women at older ages in India and dominate all other sources of excess female mortality. Finally, congenital deaths at infancy, as well as Injuries, account for a suspiciously large total of excess female deaths in India. These excess deaths easily outnumber maternal mortality.

In sub-Saharan Africa, missing girls also die prematurely from preventable diseases: malaria is a primary killer. As to India, maternal deaths are also important. But *the* dominant source of missing women in sub-Saharan Africa is HIV/AIDS. It accounts for well over a third of excess female deaths in the region. (That said, sub-Saharan African percentages of missing women are still comparable to those in India and China even if the excess female deaths from HIV/AIDS is entirely ignored.)

In China, by contrast, the dominant source of missing females is prenatal. That said, there are excess female deaths in childhood which are due to respiratory and perinatal causes. To us, these are warning signs that active female discrimination in China possibly stretches beyond the prenatal. Indeed, a large chunk of missing women in China, as well as in India, are found after the age of 45. In China, these excess deaths from Group 2 diseases account for close to 40% of the flow of all missing women. The corresponding figure for India is also 40%. These numbers point to the importance of studying the conditions of elderly women in India and China.

As a final note, we observe some similarities between age-specific percentages of missing women in the historical United States (ca. 1900) and India or sub-Saharan Africa today.

Our exercise cannot disentangle the role of direct gender discrimination from other factors (biological, social, environmental, behavioural, or economic) in explaining "missing women". But it allows us to seek out potential pathways of influence, and to assess the comparative contributions made by various categories to overall excess female mortality. In this context, observe that "category" need not be circumscribed by just age and disease. For instance, inspired by the work of Das Gupta (1987), Muhuri and Preston (1991), and others on family composition and female mortality, we could—in principle—assess the contribution of the category "girls aged 0–5 with female siblings" to overall excess female mortality. We could compare this category with something entirely different, such as "Injuries". Because our approach can easily put a variety of categories into one unified framework, it may have the potential to inform future comparative research in these areas.

APPENDIX A. DESCRIPTION OF DATA SOURCES

Throughout, we use data from the World Health Organization (WHO). The WHO uses a variety of sources to obtain estimates of death rates.[59] For developed countries, data on population numbers, births, and deaths come from vital registration data. For our key regions of interest, reliable vital registration systems are generally incomplete. At times they are entirely non-existent (particularly in parts of sub-Saharan Africa).

In the absence of complete vital registration data, the WHO combines the most recent census and survey materials together with demographic techniques to compute their estimates. For China, the availability of the 2000 census (which collected data on all deaths between 1999 and 2000 for each household) made direct estimates of age-specific death

59. Consult Mathers *et al*. (2004) for a detailed description of data sources and methods for these estimates.

rates possible.[60] Using demographic techniques, it is possible to estimate the incidence of under-reporting for both males and females and the data we use is adjusted to correct for this.[61] For India, separate mortality and recording systems for rural and urban areas were used to estimate death rates by age and sex for rural and urban areas and these were added to obtain national death rates. The all-cause mortality rates were derived from a time series analysis of age-specific death rates from the Sample Registration System after correcting them for under-registration.[62]

Relative to China and India, the task of estimating reliable mortality rates for sub-Saharan Africa is more challenging. Recent vital registration data are only available for 20% of the countries in sub-Saharan Africa. Otherwise, the main data sources used include the Demographic Health Surveys (which cover 80% of countries in sub-Saharan Africa) as well as census data (available for 73% of countries).[63] Other sources include the World Fertility Surveys (a predecessor of the DHS surveys), the Multiple Indicator Cluster Surveys (collected by UNICEF), and National Integrated Houshold Surveys (akin to the Living Standard Measurement Surveys collected by the World Bank). Using all of the data at hand, together with regression techniques and a set of roughly 2000 life tables judged to be of good quality, the WHO computed estimates for mortality rates (excluding HIV/AIDS and war deaths) by age and gender for all sub-Saharan African countries. HIV/AIDS deaths and war deaths were then added to total mortality rates where necessary.

The reliability of these estimates, for sub-Saharan Africa in particular, has been challenged (Cooper *et al*., 1998). But we have no other data at our disposal. Because our aim is to compare estimates of missing women across different regions of the world, the efforts made by the WHO to compute comparable mortality rates worldwide is the best we can depend upon. Where possible, we compare the estimates we use here to alternative data sources. In general, we find no significant disparities. This is probably the case because the WHO has already incorporated any alternative micro-level survey data that we have access to (for example, Census information for India and China, and DHS surveys for sub-Saharan Africa), to derive their estimates.

The GBD used numerous data sources and epidemiological models to estimate the first comprehensive worldwide cause-of-death patterns in 14 age–sex groups for over 130 important diseases.[64] The estimates reflect all of the information currently available to the WHO.[65] On a routine basis, member states are required to submit the latest cause-of-death data from their vital registration resources. In the absence of a complete and accurate vital registration system, countries are requested to submit all other reliable sources. For these countries (primarily in the developing world), cause-of-death data have been carefully analysed to take into account incomplete coverage of vital registration data and the likely differences in cause-of-death patterns that would be expected in the uncovered and often poorer sub-populations. It is important to be aware of the limitations of this dataset.

Cause-specific mortality data for China is available from two sources: the sample vital registration system, monitored by the Ministry of Health; and the Disease Surveillance Point System, monitored by the Chinese Center for Disease Control. The population covered by these two systems is approximately 140 million. Both systems

60. Aside from vital registration data for 1987–2000, the entire set of sources used for China include the 1990 and 2000 census; Disease Surveillance Points 1991–1990; Fertility Sampling Survey 1992; National Survey on Fertility and Birth Control 1988; Female Fertility in China: Population Survey 1982; Population Sample Survey 1987, 1990–1994, 1995, and 1996–1998; Child and Maternal Surveillance System 1991–1998.

61. Demographic techniques to assess the completeness of recorded mortality data are based on certain assumptions regarding the stability of population growth rates and migration. Refer to Hill (2003) for a summary of the methods used to estimate mortality rates in developing countries.

62. Apart from vital registration data from 1990 to 1999, other sources used include Census of India 1981, 1991; National Family Planning Survey 1970; Second All India Planning Survey 1980; Survey on Infant and Child Mortality 1979; National Family Health Survey 1992, 2000.

63. The DHS surveys collect complete sibling histories from repsondents. Hill and Trussell (1991) developed a method to estimate sex- and age-specific death rates from this information. Inter-censal survival data can be used together with demographic methods such as the "growth-balanced" technique to compute sex and age-specific death rates. Refer to Gakidou *et al*. (2004) for more discussion.

64. Worldwide comparability of cause-of-death data has been made possible through the development and revisions of the International Statistical Classification of Diseases (ICD), adopted in the early 1990s. To be sure, accuracy in diagnosing causes of death varies by country. To produce unbiased estimates of cause-specific death rates and to maximize comparability across countries, deaths coded to ill-defined categories are redistributed *pro rata* across all causes excluding injuries. Correction algorithms are also applied to resolve problems of miscoding. The cause categories used for the GBD study follow the principles of ICD that each death is categorically attributed to one underlying cause.

65. The GBD study uses the all-cause number of deaths, by age and sex, to provide an "envelope" which constrains individual disease and injury estimates of deaths. Therefore, the sum of deaths from all specific causes for any age–sex group must sum to the estimates of the total number of deaths for that age–sex group.

have sample sites which are classified into urban and rural and also into different socioeconomic strata. Age-specific mortality rates for specific conditions and for each stratum of the population are estimated and these are then aggregated up proportionately, corrected for under-registration, and adjusted with information from WHO technical programmes on specific cause mortality.

For India, cause-specific patterns of mortality were based on the Medical Certificate of Cause of Death for urban areas and the Annual Survey of Causes of Death for rural areas. Additional sources from large-scale verbal autopsy studies were used to implement an algorithm to redistribute ill-defined deaths to specific causes. The cause-specific mortality estimates at the aggregated national level were then adjusted from WHO technical programmes on specific cause mortality.

For countries without vital registration data on cause of death (and most of sub-Saharan Africa comes under this heading), the WHO estimated cause-of-death models using maximum likelihood techniques. Estimates of death rates across diseases were based on estimated total mortality rates and average *per capita* income. Regional model patterns of specific causes of death within each cause group have been constructed from vital registration data from neighbouring countries with similar patterns of mortality and income. Specific causes are further adjusted on the basis of epidemiological evidence from registries, verbal autopsy studies, disease surveillance systems, and analysis from WHO technical programmes. For sub-Saharan Africa in particular, a regional model pattern of specific causes of deaths was based on vital registration data from urban and rural South Africa.[66] The weights used for each country were determined by overall mortality death rates and income levels of those countries.

There are other hurdles. In general, deaths resulting from war are not systematically included in causes of death from the vital registration system. Likewise, deaths due to AIDS and drug use are typically undercounted. In most cases, adjustments for deaths due to these causes have been made using other sources. For instance, country-specific estimates of HIV and AIDS mortality have been developed by the joint United Nations Programme on HIV/AIDS and WHO and revised periodically. Country-specific estimates of war deaths and corresponding uncertainty ranges were obtained from a variety of published and unpublished war mortality databases.[67] Epidemiological estimates for key causes in developing countries are also taken into account. For example, for Africa, country-specific estimates of malaria mortality were based on an analysis by Snow *et al*. (1999) and updated using the most recent geographical distributions of risks in MARA (Mapping Malaria Risk in Africa). For maternal mortality, various surveys such as the DHS were used together with epidemiological models.

Faced with the numerous and necessary interpolations as well as the reliance on a variety of data sources (none of which was systematically created to seamlessly work with the others), it is tempting to dismiss the GBD results as uninformative at best, or even misleading. On balance, while we recognize the many limitations of the data and have taken pains to point them out here, however briefly, this is not our conclusion. *Of course*, caution is required when inferring comparability of national disease burden assessments across countries. This is especially true of estimates where there is no functioning vital registration system for cause of death. Judged by the demanding standards of detailed micro-surveys, the GBD is obviously not good enough. However, judged by the standards of data used in macroeconomic cross-country regressions, it is probably pretty good. In any case, it is the best data we have.

APPENDIX B. OTHER REFERENCE BENCHMARKS

The paper uses developed countries as a reference group. Table B1 recomputes our estimates of missing women in our paper using Latin America and the Caribbean as the reference group. For each region, the first of the two columns reproduces those in Table 3 from the paper (with developed-country benchmarks), while the second column uses the Latin American/Caribbean benchmarks.

Before interpreting these estimates it should be noted that, relative to developed countries, 0.06% of the female population is missing in Latin America and the Caribbean and the majority are of age 60 and older. These missing women in Latin America and the Caribbean correspond roughly to the decrease in missing women we observe if we instead use this region as the reference group compared to our estimates in Table 3. The estimates for China decrease the most, and the percentage of missing women in China older than 15 is now 40%, compared to 55% in the main calculations. For India, the percentage of missing women older than 15 falls from 66 to 59%. The estimates

66. The exception is Zimbabwe, which also has data on cause-specific death rates from the vital registration system. For some countries, other sources were also available such as hospital mortality data, deaths certified by medical personnel, and verbal autopsy reports. This was the case for the Ivory Coast, Ghana, Kenya, Madagascar, and Senegal.

67. Deaths due to landmines and unexploded ordnance were estimated separately.

TABLE B1

Excess female deaths by age (in 000s), using Latin America/Caribbean as Benchmark

	India		China		ssAfrica	
Age group	DC	LAC	DC	LAC	DC	LAC
At birth	184	184	644	644	0	0
0–1	146	146	109	109	32	32
1–4	164	150	23	19	160	127
5–9	62	60	2	2	40	37
10–14	31	28	−0	−1	30	27
15–19	77	81	−1	2	98	103
20–24	102	104	7	9	222	226
25–29	79	80	18	18	258	260
30–34	50	55	24	28	195	208
35–39	17	27	26	32	103	121
40–44	27	30	23	24	47	52
45–49	24	15	33	25	24	14
50–54	41	26	28	15	25	13
55–59	56	20	29	1	35	16
60–64	86	30	53	0	43	14
65–69	155	79	100	17	57	23
70–74	188	126	150	70	62	34
75–79	112	63	185	106	50	27
80–84	72	42	151	84	30	14
85–89	32	19	83	47	11	4
90–94	9	5	31	16	2	0
95–99	1	0	6	1	0	0
100+	0	0	1	−1	0	0
Total (mw_A)	1712	1368	1727	1267	1526	1352
% Female population	0.34	0.28	0.31	0.20	0.47	0.42

Numbers do not sum to total, or may appear identical because of rounding error.
Sources: Global Burden of Disease (2002) and Table 3.

for sub-Saharan Africa are affected the least. The main findings are qualitatively unaltered, even though (as already stated) we are uncomfortable with this approach.

Several demographers who have researched missing women argue that a better reference group is to use Coale–Demeny model life tables (see Coale, 1991; Klasen and Wink, 2003). Model life tables are primarily used to predict mortality rates when data is scarce. The Coale–Demeny tables use historical data (mainly from Europe) from as far back as before 1870 to just after the war in 1945 to predict mortality rates by age and gender at varying overall mortality and population growth rates. There are four different model life tables (North, South, East, and West) built off different regional data.

The phrase "model tables" might suggest that these are best estimates of "natural death rates" at different levels of development; however, that is not the case, as the tables are indeed built off actual data. Therefore, they incorporate existing social conditions around death, including possible gender discrimination or other (preventable) factors making for excess female mortality. In any case, when we use these model life tables as a reference group, our estimates of missing women are significantly reduced. For China, the annual total missing women is roughly 950,000 and for India it is roughly 650,000. We generally no longer have many missing women older than 60, and for some of the model life tables, we actually have missing men in these age categories. This is particularly true for sub-Saharan Africa, where the overall estimates of missing women are very small, as we have so many missing men in the older age categories. If, on the other hand, we consider only missing women aged 1–45, then we estimate roughly 400,000 missing women each year for sub-Saharan Africa.

Acknowledgements. Anderson is grateful for financial support from the Social Science Research Council of Canada and the Canadian Institute for Advanced Research. Ray's research was supported by the National Science

Foundation under grant number 0617827. We are grateful to a Co-Editor and four anonymous referees for comments and suggestions that significantly improved the paper. We thank Jean-Marie Baland, Monica Das Gupta, Patrick Francois, Parikshit Ghosh, Stephan Klasen, Ana Cecilia Lima-Fieler, Emily Oster, Anja Sautmann, and seminar participants at the University of Victoria, PRIO (Oslo), CIFAR (Toronto), Simon Fraser University, McGill University, the World Bank, and the BREAD meeting at Brown University for very useful comments.

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 EDLUND, L. (2008), "Son-biased Sex Ratios in the 2000 United States Census", *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 5681–5682.

ANDERSSON, R. and BERGSTROM, S. (1998), "Is Maternal Malnutrition Associated with a Low Sex Ratio at Birth?" *Human Biology*, **70** (6), 1101–1106.

BLOCH, F. and RAO, V. (2002), "Terror as a Bargaining Instrument: Dowry Violence in Rural India", *American Economic Review*, **92** (4), 1029–1043.

BONGAARTS, J. (1996), "Global Trends in AIDS Mortality", *Population and Development Review*, **22** (1), 21–45.

CATANIA, J., OSMOND, D., STALL, R., POLLACK, L., PAUL, J., BLOWER, S., BINSON, D., CANCHOLA, J., MILLS, T., FISHER, L., CHOI, K., PORCO, T., TURNER, C., BLAIR, J., HENNE, J., BYE, L. and COATES, T. (2001), "The Continuing HIV Epidemic Among Men Who Have Sex with Men", *American Journal of Public Health*, **91**, 907–914.

CHAHNAZARIAN, A. (1988), "Determinants of the Sex Ratio at Birth: Review of Recent Literature", *Social Biology*, **35** (34), 214–235.

CHEN, M. and DRÈZE, J. (1992), *Widows and Well-being in Rural North India* (Mimeo, London School of Economics).

CHENINE, A.-L., SHAI-KOBILER, E., STEELE, L., ONG, H., AUGOSTINI, P., SONG, R., LEE, S., AUTISSIER, P., RUPRECHT, R. and SECOR, W. E. (2008), "Acute Schistosoma mansoni Infection Increases Susceptibility to Systemic SHIV Clade C Infection in Rhesus Macaques after Mucosal Virus Exposure", *PLoS Neglected Tropical Diseases*, **2** (7), e265.

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), 517–523.

COALE, A., DEMENY, P. and VAUGHAN, B. (1983), *Regional Model Life Tables and Stable Populations* (Princeton, NJ: Princeton University Press).

COOPER, R., OSOTIMEHIN, B., KAUFMAN, J. and FORRESTER, T. (1998), "Disease Burden in Sub-Saharan Africa: What Should We Conclude in the Absence of Data?" *Lancet*, **351**, 208–210.

DAS GUPTA, M. (1987), "Selective Discrimination Against Female Children in Rural Punjab, India", *Population and Development Review*, **13**, 77–100.

DAS GUPTA, M. (2005), "Explaining Asia's 'Missing Women': A New Look at the Data", *Population and Development Review*, **31** (3), 529–535.

DEATON, A. (1989), "Looking for BoyGirl Discrimination in Household Expenditure Data", *World Bank Economic Review*, **3**, 183–210.

DHANDAPANY, P., SADAYAPPAN, S., XUE, Y., POWELL, G., RANI, D., NALLARI, P., RAI, T., KHULLAR, M., SOARES, P., BAHL, A., THARKAN, J., VAIDEESWAR, P., RATHINAVEL, A., NARASIMHAN, C., AYAPATI, D., AYUB, Q., MEHDI, S., OPPENHEIMER, S., RICHARDS, M., PRICE, A., PATTERSON, N., REICH, D., SINGH, L., TYLER-SMITH, C. and THANGARAJ, K. (2009), "A Common MYBPC3 (Cardiac Myosin Binding Protein C) Variant Associated with Cardiomyopathies in South Asia", *Nature Genetics*, **41**, 187–191.

DRÈZE, J. (1990), *Widows in Rural India* (Mimeo, London School of Economics).

DUBUC, S. and COLEMAN, D. (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.

DUNKLE, K., JEWKES, R., BROWN, H., GRAY, G., MCINTYRE, J. and HARLOW, S. (2004), "Gender-based Violence, Relationship Power, and Risk of HIV Infection in Women Attending Antenatal Clinics in South Africa", *Lancet*, **363**, 1415–1421.

EBENSTEIN, A. (2007), *Fertility Choices and Sex Selection in Asia: Analysis and Policy* (Mimeo, Department of Economics, Berkeley).

GAKIDOU, E., HOGAN, M. and LOPEZ, A. D. (2004), Adult Mortality: Time for a Reappraisal, *International Journal of Epidemiology*, **33**, 710–717.

GARENNE, M. (2002), "Sex Ratios at Birth in African Populations: A Review of Survey Data", *Human Biology*, **74**, 889–900.

GARENNE, M. (2003), "Sex Differences in Health Indicators Among Children in African DHS Surveys", *Journal of Biosocial Science*, **35**, 601–614.
GARENNE, M. (2004), "Sex Ratios at Birth in Populations of Eastern and Southern Africa", *South African Journal of Demography*, **9**, 91–96.
GARG, A. and MORDUCH, J. (1998), "Sibling Rivalry and the Gender Gap: Evidence from Child Health Outcomes in Ghana", *Journal of Population Economics*, **11** (4), 471–493.
GOODREAU, S. and GOLDEN, M. (2007), "Biological and Demographic Causes of High HIV and Sexually Transmitted Disease Prevalence in Men Who Have Sex with Men", *Sexually Transmitted Infections*, **83**, 458–462.
GUPTA, G. (2002), "How Men's Power Over Women Fuels the HIV Epidemic", *British Medical Journal*, **321**, 183–184.
HILL, K. (2003), *Adult Mortality in the Developing World; What We Know and How We Know It* (Mimeo, Population Division, Department of Economic and Social Affairs, United Nations Secretariat).
HILL, K. and TRUSSELL, J. (1991), "Further Developments in Indirect Mortality Estimation", *Population Studies*, **45**, 455–472.
JAMES, W. H. (1987), "The Human Sex Ratio. Part I: A Review of the Literature", *Human Biology*, **59**, 721–752.
JHA, P., KUMAR, R., VASA, P., DHINGRA, N., THIRUCHELVAM, D. and MOINEDDIN, R. (2006), "Low Male-to-female Sex Ratio of Children Born in India: National Survey of 1.1 Million Households", *Lancet*, **367**, 211–218.
JOSHI, R., CARDONA, M., IYENGAR, S., SUKUMAR, A., RAJU, C., RAJU, K., RAJU, K., REDDY, K., LOPEZ, A. and NEAL, B. (2006), "Chronic Diseases Now a Leading Cause of Death in Rural India—Mortality Data from the Andhra Pradesh Rural Health Initiative", *International Journal of Epidemiology*, **35**, 1522–1529.
JUNHONG, C. (2001), "Prenatal Sex Determination and Sex-selective Abortion in Rural Central China", *Population and Development Review*, **27**, 259–281.
KHOURY, M., ERICKSON, D. and JAMES, L. (1984), "Paternal Effects on the Human Sex Ratio at Birth: Evidence from Interracial Crosses", *American Journal of Human Genetics*, **36**, 1103–1111.
KLASEN, S. (1996), "Nutrition, Health, and Mortality in Sub Saharan Africa: Is There a Gender Bias?", *Journal of Development Studies*, **32**, 913–933.
KLASEN, S. (1998), "Marriage, Bargaining, and Intrahousehold Resource Allocation: Excess Female Mortality Among Adults in Rural Germany (1740–1860)", *Journal of Economic History*, **58**, 432–467.
KLASEN, S. and WINK, C. (2003), "'Missing Women': Revisiting the Debate", *Feminist Economics*, **9** (23), 263–299.
KOCHAR, A. (1999), "Evaluating Familial Support for the Elderly: The Intrahousehold Allocation of Medical Expenditures in Rural Pakistan", *Economic Development and Cultural Change*, **47**, 620–656.
LERCHL, A. (1998), "Seasonality of Sex Ratio in Germany", *Human Reproduction*, **13** (5), 1401–1402.
LIN, M.-J., LIU, J.-T. and QIAN, N. (2007), *Missing Women More and More: The Effect of Access to Abortion on Sex Ratios at Birth and Excess Female Mortality* (Mimeo, Department of Economics, Brown University).
LIN, M.-J. and LUOH, M.-C. (2008), "Can Hepatitis B Mothers Account for the Number of Missing Women? Evidence from Three Million Newborns in Taiwan", *American Economic Review*, **98**, 2259–2273.
MACKENZIE, C., LOCKRIDGE, A. and KEITH, M. (2005), "Declining Sex Ratio in a First Nation Community", *Environmental Health Perspectives*, **1131** (10), 1295–1298.
MAMAN, S., CAMPBELL, J., SWEAT, M. and GIELEN, A. (2000), "The Intersections of HIV and Violence: Directions for Future Research and Interventions", *Social Science & Medicine*, **50**, 459–478.
MATHERS, C., BERNARD, C., MOESGAARD IBURG, K., INOUE, M., MA FAT, D., SHIBUYA, K., STEIN, C., TOMIJIMA, N. and XU, H. (2004), "Global Burden of Disease in 2002: Data Sources, Methods, and Results" (Global Programme on Evidence for Health Policy Discussion Paper No. 54, World Health Organization).
MATHEWS, T. and HAMILTON, B. (2005), "Trend Analysis of the Sex Ratio at Birth in the United States", *National Vital Statistics Reports*, **53** (20), 1–17.
McMAHAN, C. (1951), "An Empirical Test of Three Hypotheses Concerning the Human Sex Ratio at Birth in the United States, 1915–1948", *The Millbank Memorial Fund Quarterly*, **29** (3), 273–293.
MENSKI, W. (ed.) (1998), *South Asians and the Dowry Problem* (New Delhi: Vistaar Publications).
MORTON, N., CHUNG, C. and MI, M. (1967), "Genetics of Interracial Crosses in Hawaii", *Monograph in Human Genetics*, vol. 3 (New York: S. Karger).
MUHURI, P. and PRESTON, S. (1991), "Effects of Family Composition on Mortality Differentials by Sex Among Children in Matlab, Bangladesh", *Population and Development Review*, **17**, 415–434.
MYERS, R. (1947), "Effect of War on the Sex Ratio at Birth", *American Sociological Review*, **12** (1), 40–43.
NATHANSON, C. A. (1984), "Sex Differences in Mortality", *Annual Review of Sociology*, **10**, 191–213.
NAVARA, K. J. (2009), "Humans at Tropical Latitudes Produce More Females", *Biology Letters*, **5**, 524–527.
NIKIFOROV, S. and MAMAEV, V. (1998), "The Development of Sex Differences in Cardiovascular Disease Mortality: A Historical Perspective", *American Journal of Public Health*, **88** (9), 1348–1353.

OMRAN, A. (1971), "The Epidemiological Transition: A Theory of the Epidemiology of Population Change", *Milbank Quarterly*, **49**, 509–537.
OSTER, E. (2005), "Hepatitis B and the Case of the Missing Women", *Journal of Political Economy*, **115** (6), 1163–1216.
OSTER, E. (2009), "Proximate Sources of Population Sex Imbalance in India", *Demography*, **46**, 325–340.
OSTER, E., CHEN, G., YU, X. and LIN, W. (2008), "Hepatitis B Does not Explain Male-biased Sex Ratios in China" (Mimeo, Department of Economics, University of Chicago).
PHILLIPS, M. R., LI, X. and ZHANG, Y. (2002), "Suicide Rates in China, 1995–99", *Lancet*, **359**, 835–840.
RONSMANS, C. and GRAHAM, W. (2006), "Maternal Mortality: Who, When, Where and Why", *Lancet*, **368**, 1189–1200.
SECOR, W. E., KARANJA, D. and COLLEY, D. (2004), "Interactions Between Schistosomiasis and Human Immunodeficiency Virus in Western Kenya", *Memórias do Instituto Oswaldo Cruz*, **99** (Suppl. 1), 93–95.
SEN, A. (1990), "More Than 100 Million Women Are Missing", *The New York Review of Books*, **37** (20), 1990–.
SEN, A. (1992), "Missing Women", *British Medical Journal*, **304**, 587–588.
SILVERMAN, J., DECKER, M., GUPTA, J., KAPUR, N. and RAJ, A. (2007), "Violence Against Wives, Sexual Risk and Sexually Transmitted Infection Among Bangladeshi Men", *Sexually Transmitted Infections*, **83**, 211–215.
SNOW, R., DEICHMANN, C. and MARSH, K. (1999), "Estimating Mortality, Morbidity, and Disability Due to Malaria Among Africa's Non-pregnant Population", *Bulletin of the World Health Organization*, **77** (8), 624–640.
SUBRAMANIAN, S. and DEATON, A. (1991), "Gender Effects in Indian Consumption Patterns", *Sarvekshana*, **14**, 1–12.
SUDHA, S. and RAJAN, S. I. (1999), "Female Demographic Disadvantage in India 1981–1991: Sex Selective Abortion and Female Infanticide", *Development and Change*, **30**, 585–618.
TARVER, J. and LEE, C. (1968), "Sex Ratio of Registered Live Births in the United States, 1942–63", *Demography*, **5**, 374–381.
ZENG, Y., TU, P., GU, B., XU, Y., LI, B. and LI, Y. (1993), "Causes and Implications of the Recent Increase in the Reported Sex Ratio at Birth in China", *Population and Development Review*, **19**, 283–302.