Missing Women: Age and Disease

Siwan Anderson
University of British Columbia
and

Debraj Ray
New York University

June 2008

Abstract

Relative to developed countries and some parts of the developing world, most notably sub-Saharan Africa, there are far less women than men in India and China. It has been argued that as many as a hundred million women could be missing. The possibility of gender bias at birth and the mistreatment of young girls are widely regarded as key explanations. While we do not dispute the existence of severe gender bias, our computations of the number of missing women at different ages and by cause of death 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; and (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.

Ray's research was supported by the National Science Foundation under grant number 0617827. Anderson is grateful for financial support from the Social Science Research Council of Canada and the Canadian Institute for Advanced Research. We thank Jean-Marie Baland, Patrick Francois, Parikshit Ghosh, Ana Cecilia Lima-Fieler, Emily Oster, Anja Sautmann and seminar participants at the University of Victoria, PRIO (Oslo) and the BREAD meeting at Brown University for very useful comments.

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. It is well known that on average males tend to outnumber females at birth, but that this imbalance begins to redress itself soon after. It is certainly true that in developed countries, men die faster than women at every age group after birth (see Coale (1991) for a discussion and the Regional Model Life Tables in Coale, Demeny and Vaughan (1983)). The combined effect is — or "should be" — a roughly equal proportion of men and women in the population as a whole, provided that males and females are provided with unbiased health and nutritional environments.[1] That is certainly not the case in large parts of Asia: in particular, in India and China, the overall ratio of males to females is around 1.06.

In 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". Later in his article, he attempts to explain why this might be the case, arguing that female participation in the labor force is relatively high in sub-Saharan Africa. Indeed, as Sen observes, such participation is at the top of the list among all the developing regions with substantial demographic imbalances in gender.

There is little doubt that several cultural and economic factors that systematically discriminate against women lie at the heart of this problem. A particular focus of attention has been the sex ratio at birth, which may be indicative of practices such as selective abortion following pre-natal sex detection (see, for instance, Junhong (1991), Sudha and Rajan (1999), Ebenstein (2007) and Lin *et al.* (2007)). 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. Likewise, 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.

A second area of focus is early childhood, where it has been argued that young girls are systematically less cared for (see, e.g., Deaton (1989), Subramanian and Deaton (1991), and Garg and Morduch (1998)).[2] Some or much of this may manifest itself in higher death rates for girls, and to the extent that female births (followed by neglect or infanticide) go unreported, it may manifest itself in skewed ratios at birth as well. 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. It is not that the later-age travails of women are entirely ignored (see, for instance, Drèze (1990), Drèze and Chen (1992), and Kochar (1999)), but the accusing finger seems to have been predominantly pointed at the pre-natal and infant/early childhood stages. Das Gupta

[1] The ratio will vary a bit depending on life expectancy and the rate of 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 favors men (1.022) when population is growing at 2% per year, and women (0.997) in stationary populations.

[2] Oster (2008) examines the role of vaccinations and malnutrition in explaining excess female mortality before the age of 5 in India.

(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" (emphasis ours).

The exercise in this paper is extremely simple. In contrast to the admittedly important focus on childhood or pre-natal discrimination, we study how missing women in different countries are distributed across all age groups (and later by age and disease groups). The methodology we employ is in the spirit of the Sen-Coale exercises. For each category — age, and later age and disease — we obtain an "unbiased" death rate for females, one that would obtain 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.[3] We subtract this unbiased rate from the actual death rate for females, and then multiply by the population of females in that category. Such a benchmark raises questions. It is neutral on scale: after all, high-death countries assuredly have both missing women and men. But we ignore these, and only count the number of "net" missing women. We return to this issue of benchmarking later in the paper.

Our exercise has the modest aim of simply assessing how missing women are distributed across age groups. But it leads to a bit more than that. First, although the *overall* ratios of males to females in India and China are similar (they are both around 1.06), the experiences of China and India are quite distinct. True to the emphasis in the literature, a large percentage of the missing women in China are to be found before birth and in infancy. (We estimate that around 37–45% of the missing women in China are due to pre-natal factors alone.) But those for India are more evenly distributed across the different age groups. Pre-natal factors account for around 11%, and if we add up all the missing women up to age 15, we only get to less than a third of the total.

Indeed, both India and China exhibit large fractions of missing women in older (post-50) age groups. As we shall also see, these estimates easily dominate the corresponding numbers for maternal mortality. Second, and in stark contrast to the previous literature, we find that as a proportion of the female population, the incidence of missing women is actually largest in sub-Saharan Africa, *despite the fact that their overall male to female ratio is slightly 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"[4] 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 our Chinese and Indian estimates of 1.73m and 1.71m respectively. Expressed as a fraction of the population, the sub-Saharan numbers are significantly higher than their Chinese and Indian counterparts.

How does one reconcile these findings with the fact that sub-Saharan Africa has such a balanced overall sex ratio? The answer is simple: the sex ratio at birth in sub-Saharan Africa is equal to approximately 1.03, much lower than for developed countries as a whole (which

[3]These are the so-called Established Market Economies as defined by the World Bank.

[4]It is possible to convert our flow estimates into stocks that would compare directly with the numbers obtained by Sen, Coale, Klasen and others. This will require making the assumption that the observed age distribution in these populations is the limit distribution. We are not entirely comfortable with such an exercise, especially given the relatively recent nature of the HIV/AIDS epidemic in sub-Saharan Africa. A more nuanced calculation is left for later research.

is 1.06). There appear to be genetic differences that determine this ratio. 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.

The presence of a female deficit in many age groups, not just at birth and infancy, suggests that the epidemiological transition may "explain" a large proportion of missing women. The epidemiological transition, as is well known, refers to progressive changes in the causal composition of mortality, with infectious diseases giving way to chronic and degenerative ailments as the leading causes of death. Briefly, infectious disease and nutritional or reproductive health problems predominate in poorer, high-mortality populations, while chronic and degenerative ailments predominate in low-mortality populations. Infectious diseases and undernutrition do not discriminate across gender, and in some obvious cases (say reproductive mortality) contribute to push death rates for women above those for males.[5] Aggregative age-by-age calculations that do not control for the effects of disease composition might then record a number of "missing women". But it isn't clear at all that we should label such women as "missing" in the first place! They would certainly not be missing due to any want of "similar care".

In the second part of the paper, we turn to this issue and ask the following question: what proportion of the missing women are missing because the disease composition is different between developed and developing countries? And what proportion is missing *disease by disease*? Our findings surprised us. As best as we can understand within the scope of existing data, disease-by-disease effects account for an overwhelming fraction of all missing women. In India, the number is 1.64m out of the aggregate estimate of 1.71m, in sub-Saharan Africa 1.39m out of a total of 1.53m, and in China 1.59m out of 1.73m. It does appear that at all ages, women do die from lack of similar care, though "similar care" continues to be a loaded term that we will have to discuss more carefully below.

Our decomposition of missing women by age and by disease also permits us to form some idea of just which diseases are responsible. In general, at younger ages the bulk of missing females come from the so-called Group 1 diseases: infectious, perinatal and nutritional. At older ages the bulk of them come from Group 2 noncommunicable diseases. But there are differences across the three regions.

In India, respiratory and perinatal deaths are the main source of excess group 1 female deaths. Indeed, maternal mortality is also important but is of the same order of magnitude as the other two categories. But the main source (by a long way) of excess female deaths is cardiovascular. This appears to dominate all other sources of excess female mortality, and generates numbers that easily outstrip missing females at birth. That stated, an important and rather ominous category for excess female deaths in India is "Injuries". There are more missing women here than maternal mortality (for example) can account for.

In sub-Saharan Africa, much of the female deficit is to be found at younger ages. Malaria is an important component of Group 1 disease that accounts for excess female mortality. Maternal mortality is also extremely important, significantly more so than in India. But *the* dominant source of missing women is HIV/AIDS. We estimate that there are over 600,000 excess

[5]Acemoglu and Johnson (2007) examine the effects of the epidemiological transition on life expectancy today.

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 (see Section 4.2.5). For instance, the female deficit in cardiovascular disease or malaria is also large in sub-Saharan Africa.

In China, Group 1 diseases do not play as large a role. The main excess female deaths in this category, as in India, are perinatal and respiratory. The large female deficit in Group 2 stems, once again, from cardiovascular disease, but respiratory disease is of comparable importance in older ages.

We believe these findings are important because they don't just say that there are missing women, but also tell us where they are to be found (in age-disease space). The results also suggest that some sort of lack in "similar care" is at work, rather than the relatively benign discrepancies generated by a change in the disease mix.

At the same time, our findings also raise questions about just what "similar care" (or the lack of it) means. As described earlier, in the spirit of the Sen-Coale exercise, we simply use the relative death rates of males and females in developed countries as reference points.[6] Does this mean that if there are missing women due to HIV/AIDS, that women literally do not receive similar care in *treatment*? Certainly not: much of the excess is also due to relative differences in cultural and sexual norms, which give rise to an entirely different route of transmission (which, in the case of HIV/AIDS in sub-Saharan Africa, turns out to be mainly heterosexual). Is this a "lack of similar care"? In some broad sense, yes, but not so in the narrow sense of treatment. Similarly, in the case of cardiovascular illness, excess female deaths may refer to treatment, but also to differential incidence: "similar care" has a broader connotation here as well. We discuss this issue in detail in Section 6.

More generally, our 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. In a sense, cultural norms have been largely to blame. But our estimates suggest that excess female mortality is a much more universal phenomenon than previously thought. We find that the majority of excess female deaths actually occur at older ages, when parental preferences should not be having a direct role, and also in other regions (particulary Sub-Saharan Africa) with very distinct cultural norms.

To further develop this point, we take our methodology to historical data for now-developed countries. Using early data from the United States, we compute the number of missing women in 1900. There is a remarkable congruence between these numbers and what we observe in India and China (and in sub-Saharan Africa) 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.

Our findings imply that if we want to restrict ourselves to defining missing women as the number of girls who have died due to discriminatory parental preferences, then the original estimates need to be seriously revised downwards, but also that such a computation is not at all straightforward. For example, at most 11% of the total missing women in India occur

[6]Maternal mortality is an important exception; we deal with it separately.

before birth. As a result only a relatively small proportion of the original total estimate of missing women in India can potentially be directly attributed to biased sex ratios at birth. Likewise only 23% of the total missing women in India are of childhood age (0 to 14). However, attributing this childhood female deficit directly to discriminatory *cultural* practices is made problematic by the fact that a comparable proportion of girls in this age group are *also* missing in Sub-Saharan Africa.[7]

Of course, it is entirely possible (though we doubt it) that all missing women at older ages are due to discriminatory treatment in childhood. For instance, a history of childhood malnutrition can certainly increase susceptibility to disease at older ages. Indeed, in this paper we do not attempt to disentangle the various channels which "explain" excess female mortality at different ages and diseases. This is an important research agenda.[8] At the same time, our findings take an important first step in this direction. Knowing that coronary or respiratory disease, or "injuries", or HIV/AIDS may be responsible for a female deficit allows us to look in the right places, and serves as a starting point for in-depth research aimed at understanding the underlying causes.

2. Computing Missing Women

2.1. **The Sen Counterfactual.** Numbers speak louder than ratios, and 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 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).

We've already remarked that 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. Coale (1991, Table 1) adopts a similar procedure, but 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

[7]According to our estimates, 0.25% of girls aged 0 to 14 are missing in India. The proportion missing of the same age in Sub-Saharan Africa is 0.17%.

[8]For instance, Qian (2008) studies how economic factors impinge on biased sex ratios in China.

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.

The initial procedure we adopt is a modified version of the Sen counterfactual applied to every age group. This approach allows us to identify missing women at various age categories and does not (necessarily) seek a single overall number. This procedure certainly deals with Coale's objection that a comparison of overall sex ratios ignores the role of different age compositions across regions. But quite apart from this, the description of missing women at different ages (and later, by age and disease) is central to our inquiry.

Because current sex ratios reflect deaths and births over several decades past, as well as migration flows, our procedure uses death rates by age (instead of sex ratios) for our computation of missing women in each age group. The exception to this is the computation of excess prenatal female deaths where we have to rely on sex ratios at birth. A concern here is the variation in the sex ratio at birth. Ideally, one would like to examine the sex ratio at birth among the progeny of Indians, Chinese, sub-Saharan Africans in situations where pre-natal selection is unavailable. Unfortunately, for reasons we discuss below, this is an extremely difficult task. One feature that does stand out, and which we incorporate in our analysis, is the fact that there is a significant difference in the sex ratio *at birth* for sub-Saharan Africa as a whole, compared to other groups.[9] The reference sex ratio at birth we employ in each case is discussed in more detail below.

Next, we extend the age decomposition of missing women to disease, thereby obtaining estimates by age and disease. One reason for this is that the disease mix is clearly different across developed and developing countries. It is, therefore, unclear that all differences in death rates by gender are to be attributed to the lack of "similar care" for males and females. An entirely nondiscriminatory world in which infectious disease is dominant may still see higher death rates for females relative to men, compared to another nondiscriminatory world in which all deaths are due to, say, heart disease. An aggregative approach (or even a decomposition by age alone) would label them as "missing". That interpretation would fail to acknowledge the changing disease mix or what is commonly referred to as the epidemiological transition, just as an aggregative calculation will fail to account for the changing age composition. The next few subsections take up these matters.

2.2. **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. With this in mind, let $d^m(a)$ and $d^w(a)$ represent the rate of death of men and women respectively at age a in the country of interest. Use the label $\widehat{\ }$ to denote these variables for a reference country. The *unbiased 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 unbiased death

[9]There is also significant variation within sub-Saharan Africa, as we observe below.

rates for women, weighted by the number of women in that age group:

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

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

As in Sen's counterfactual, the implicit presumption is that there are two differences between our country 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. The expression (2) implies that the former effect is not included in our computation of missing women; it is only the latter.[10]

2.3. **Missing Women by Age and Disease.** In preparation for studying the causes of death, we introduce 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 unbiased death rate of women at age a from disease k in the country of interest is entirely analogous to (1), and 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

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

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

As discussed in the Introduction, there is no presumption that this benchmark captures "similar care" in the narrow sense of treatment. There are social, cultural and economic differences that lead to varying degrees of incidence of a disease, as well as treatment. It may even be the case that there are excess *men* who are missing in *developed* countries, at least for some diseases or in some age groups. But this fundamental relativity is something that we cannot get away from, and we have to benchmark the story somewhere.

2.4. **Sex Ratios at Birth.** To the estimates of missing women at varying ages (and for different diseases) 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 simply not available.

Coale (1991) used a reference sex ratio at birth of 1.059 for all groups. This is problematic for a simple reason: there appears to be substantial variation in the sex ratio at birth across race

[10] Still, as Jean-Marie Baland has pointed out to us, it is unclear what the "unbiased" death rate for women should be. Our formula embodies the premise that the death rate for men in the country of interest is somehow the "right" death rate, given the overall economic conditions, and that the female death rate is distorted. It is possible that both death rates are distorted: female upwards, male downwards. (This would happen, for instance, if the total budget for health care is somehow fixed at each stage of development.) In that case the unbiased female death rate would be higher. More research needs to be done on conceptualizing these benchmarks.

Nationality/Ethnicity	Sex Ratio at Birth
White	**1.054**
Black	**1.030**
sub-Saharan African	**1.035**
American Indian	1.031
Japanese	1.055
Hawaiian	1.054
Chinese	**1.074**
Filipino	1.072
Asian Indian	**1.066**
Puerto Rican	1.045
Cuban	1.054
Central and South American	1.044
Mexican	1.041

TABLE 1. Sex Ratios at Birth by Nationality/Ethnicity in the U.S. Sources and 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 (2008) for the years 1992-2004. The numbers for sub-Saharan African parents come from IPUMS United States, 2000.

and ethnicity. It is indeed true that the average sex ratio at birth for developed countries is in the range of 1.05 to 1.07 male to female births with a median equal to 1.059, but this range is nontrivial. For example, just within Europe, the average sex ratio at birth in Northern Europe is closer to 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 U.S. is around 1.07. Table 1 summarizes some of this information, which we've obtained from different sources.

Certainly, there are a number of behavioral, biological, and environmental factors which can explain part of the variation in sex ratios at birth. 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.

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.[11] Studies demonstrate that the strong racial effect persists when controlling for other factors such as parental age, birth order and parity and these latter effects, in turn, decrease in their importance (Chahnazarian (1988)).[12]

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 U.S. may well be practicing gender selection at the prenatal stage, like their fellow nationals back home (Almond and Edlund (2008), Abrevaya (2008)).[13] 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 U.S. 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 U.S. 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 1.059–1.066. 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 U.S (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 U.S. (Table 1).[14] 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 U.S. (Table 1).

---

[11]The mean sex ratio at birth between 1915-1948 is 1.059 for whites and 1.029 for blacks (McMahan (1951)); between 1942-1963, 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)).

[12]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)).

[13]See also Dubuc and Coleman (2007) for evidence from the U.K.

[14]In the light of footnote 16 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.

Turn now to 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.[15] Those for China are substantially higher. The 2000 Census places this ratio at 1.169. The U.N. Demographic Yearbook reports a somewhat lower number for 1989 (1.139).

These large ratios are to be contrasted with their counterparts from 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 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.[16]

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:

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

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 in a given period for the very same group.

We've already discussed some basic worries with the ready use of a particular reference ratio. In addition, there is some evidence that environmental factors determine the sex ratio at birth, and some correction could be useful on this account.[17] It is also likely that that the sex ratio at birth is affected by development. There is evidence that male offspring are more susceptible to death in utero than the female, and via this channel it is possible to link maternal malnutrition to a lower sex ratio at birth (Andersson and Bergstrom (1998)). With better health care and greater prevention of such deaths, 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.

It is unclear to us that making further fine-tuned corrections would be that useful in the absence of much more comprehensive information regarding the sex ratio at birth for these

[15]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, 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.

[16]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.

[17]For instance, there is evidence that the sex ratio at birth can be altered by chemical exposure: lower proportions of male offspring have been observed in populations exposed to dioxin, mercury, pesticides, PCBs, and also parental smoking (MacKenzie *et al.* (2005)). Others have connected variations in the sex ratio at birth to wars (Myers (1947)) and seasons (Lerchl (1998)). There has been a recent debate on the effect of Hepatitis B on sex ratios (Oster 2005, Lin and Luoh 2008, and Oster *et. al.* 2008)

groups in developed countries. We therefore stick with the best available comparisons that we have. As a point of reference, though, it may be useful to put on record 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, but nevertheless provides some sense of the magnitude of any potential correction.

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

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

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 both for 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:

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

(Observe that in both cases, we include the same number at age 0, which is our estimate for 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. In fact, nowhere is the disease mix of the reference country taken into account in this procedure: all we use are the *relative* death rates for men and women in the reference country for each disease. That permits us to make the following elementary observation:

Observation **1.** *Assume that in the country of interest, the disease mix is weighted in favor 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 favor of diseases with higher relative male death rates in the country of interest. This difference may be regarded as a proxy for "missing women" due to the epidemiological transition, while the value of mw_B is a proxy for "missing women" due to lack of "similar care".

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 the epidemiological transition 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 2.2 and in (6). Throughout, we use data from the World Health Organization (WHO) and the U.N. Population Division.[18] 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 ar 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 specifice death rates possible.[19] Using demographic techniques it is possible to estimate the incidence of underreporting for both males and females and the data we use is adjusted to correct for this.[20] 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 underregistration.[21]

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).[22] 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

[18]Consult Mathers et. al. (2004) for a detailed description of data sources and methods for these estimates.

[19]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, 1990–94, 1995, and 1996–98; Child and Maternal Surveillance System 1991–1998.

[20]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.

[21]Apart from vital registration data from 1990–1999, 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.

[22]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.

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.[23]

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.

3.1. **Sex Ratios by Age.** We begin by drawing attention to how sex ratios vary by age across different countries. We will argue that India and sub-Saharan Africa share an important feature: the graph of sex ratios by age is relatively flat. However, in China and in developed countries, this graph is significantly downward sloping. In particular, India and China may have similar sex ratios overall, but they have them for very different reasons. In India, most of the missing women show up after childhood, and this is in stark contrast to the Chinese experience.

Table 2 lists sex ratios of males to females, by age group, for India, China, sub-Saharan Africa and a group of "developed regions", this last category comprising all areas of Europe plus Northern America, Australia/New Zealand and Japan. The first row in this table reports a familiar set of numbers: the overall sex ratios in China and India are similar and significantly above unity (around 1.06), while developed countries and sub-Saharan Africa also share similar sex ratios that are significantly *below* unity (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.

But matters are different in the rest of the table. The differences are best appreciated by plotting sex ratio against age for the different countries involved. Figure 1 takes a first pass at this by graphing the numbers in Table 2. Panel *A* tells us that barring the intercept term, China and the developed countries have similar graphs: sex ratios decline significantly with age. The Chinese profile — barring the ominously higher sex ratio at birth — is not that different from that of a developed country. Likewise, and in continued contrast, Panel *B* shows that India and sub Saharan Africa now seem to have more in common, in that both regions display a relatively flat sex ratio with age.[24]

[23]Refer to the discussion in Section 4.2 for more details on mortality rate estimates from these causes.

[24]To be sure, all four of these graphs are pretty flat to start with but that's because death rates are relatively low at low ages and cannot affect the sex ratio by much.

Age	Developed Countries	India	China	sub-Saharan Africa
Overall	0.964	1.065	1.058	0.985
< 1	1.055	1.060	1.107	1.018
1-4	1.053	1.065	1.115	1.016
5-9	1.051	1.071	1.115	1.014
10-14	1.050	1.075	1.104	1.012
15-19	1.049	1.083	1.084	1.007
20-24	1.044	1.085	1.067	0.997
25-29	1.031	1.102	1.052	0.983
30-34	1.027	1.102	1.041	0.973
35-39	1.024	1.099	1.050	0.965
40-44	1.013	1.096	1.082	0.955
45-49	1.000	1.084	1.057	0.945
50-54	0.991	1.032	1.081	0.930
55-59	0.970	0.984	1.081	0.903
60-64	0.934	0.951	1.060	0.874
65-69	0.879	0.932	0.996	0.848
70-74	0.792	0.912	0.902	0.813
75-79	0.657	0.875	0.758	0.760
80-84	0.553	0.828	0.616	0.703
85-89	0.433	0.800	0.462	0.609
90-94	0.328	0.784	0.276	0.478
95-99	0.240	0.761	0.151	0.328
100+	0.170	0.653	0.095	0.199

TABLE 2. SEX RATIOS BY AGE, 2000. SOURCE: *World Health Organization and U.N. Population Division.*

These diagrams are suggestive of the idea that while India/China and the developed world/sub Saharan Africa form similar pairs as far as *overall* sex ratio is concerned, the pairings are different when one looks at age *distributions*.

One important difference between India and sub Saharan Africa appears to be that India's sex ratios not only not decline, but they 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.

The flatness of sex ratio with age in India and sub-Saharan Africa, and the contrast with China, survive the use of alternative data. For instance, the Census of India estimates the sex ratio for the age group 0–4 to be 1.071 in 2001 and 1.047 in 1991. The two numbers straddle the one we use and display a large variation.[25] It is also the case that the 2000 Census of China estimates a significantly higher sex ratio for the age group 0–4 at 1.20. However,

[25]Alternative national survey data puts the 2001 estimate at 1.047 as well (U.N. Demographic Yearbook (2004)).

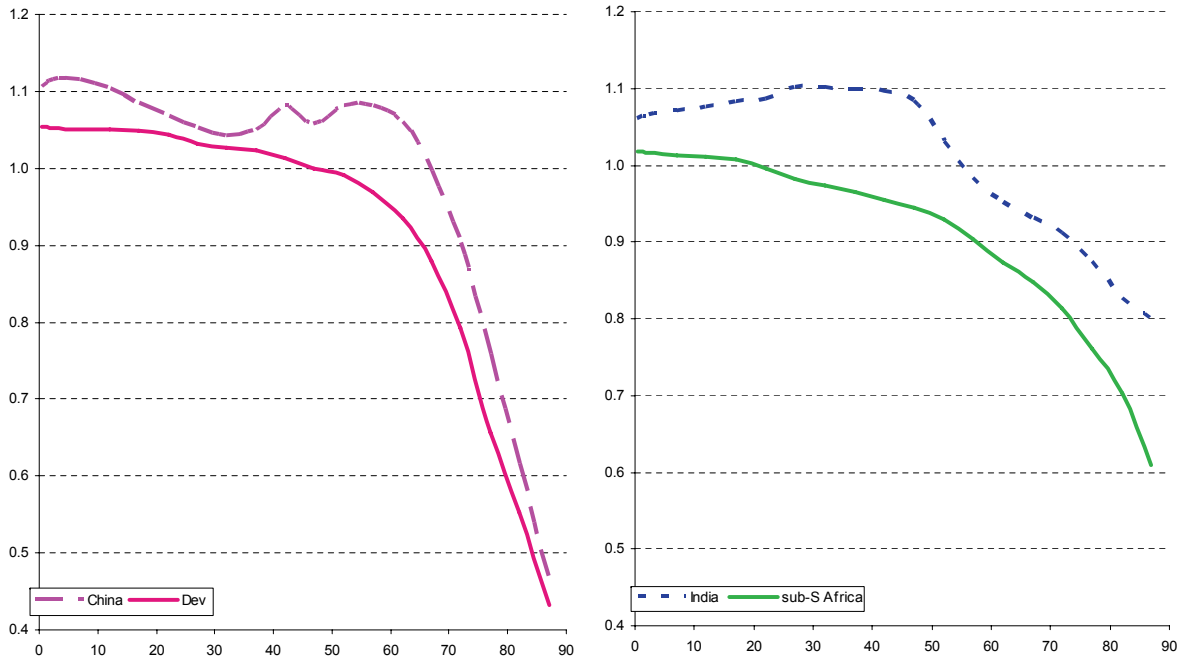

FIGURE 1. SEX RATIOS BY AGE

the key patterns found in the alternative data sources are similar to what we present here and if anything are accentuated. That is, the slope of sex ratios by age for India is always significantly flatter than that for China. Indeed, according to the 1991 Census of India, sex ratios and age are on average positively sloped until at least age 60, and moreover, the sex ratio within each age group never falls below 1.00.[26] In contrast, sex ratios by age from the 2000 China Census show a negative slope (with a very sharp decline beginning at age 60), and the lowest age-specific sex ratio is 0.35.

We do not have an alternative data source which provides estimates of sex ratios by age for the entire region of Sub-Saharan Africa. We can however, compare the pattern between sex ratios and age described above for the entire region to individual country level data. The U.N. Demographic Yearbooks provide estimates for at least 60% of Sub-Saharan countries. Consistent with the above, the overall sex ratio for these countries is on average in the range of 0.95–1.00. Similarly, the pattern between sex ratios and age is relatively flat where the range typically falls between 1.01 and 0.80. For a number of countries, sex ratios can increase with age between the ages 0 to 24 and also again at later ages (over 50).[27]

[26]Sex ratios from the 2001 Census move up and down with age, but again it is a fairly flat relationship where the sex ratio for any age group never falls below 0.95. According to alternative national survey data for India from 2001 (listed in the 2004 U.N. Demographic Yearbook) the sex ratio for any age group never falls below 0.99.

[27]For example, an increasing relationship between sex ratios and age before the age of 25 is found for Benin, Burkino Faso, Chad, Ethiopia, Ghana, Lesotho, Madagascar, Malawi, Mozambique, Nigeria, Senegal, Somalia, Sudan, and Zimbabwe.

These patterns between sex ratios and age suggest that the distribution of missing women by age varies by region and in particular that India shares commonalities with sub-Saharan Africa and important distinctions with China.

It is important to be aware, however, that current sex ratios are stocks: they represent averages of deaths and births over several past decades, and they also ignore factors such as migration. To circumvent these issues, we will look at death rates by age for these different regions. Indeed, the pairings that we've discussed will become even more pronounced.[28]

3.2. **Death Rates by Age.** Table 3 lists death rates of males and females, as well as their relative death rates, by age group, for India, China, sub-Saharan Africa and the "developed regions".

Certain things are immediately evident: for instance, sub-Saharan Africa has very high death rates overall. This is not just a reflection of HIV/AIDS; the high rates are present across the board 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. These levels will concern us later, when we form estimates of missing women by age group. What is of greater interest at present is the particularly strong form in which the earlier discussion regarding sex ratios is backed up by this table. China and the developed countries have a much higher rate of *relative* mortality for males than India and sub-Saharan Africa. The discrepancy is particularly noticeable up to middle age. Figure 2 makes this point quite dramatically by graphing the death rate data (shown in boldface in Table 3) for all four regions.

It is true that China has lower relative male mortality at all ages compared to the developed regions, but the pattern across ages is similar. Just as in developed regions, male mortality exceeds female mortality at all ages, and the excess is quite pronounced all the way up to the early 30s. We also see the familiar hump through adolescence and the 20s, with a particular (relative) surfeit of male deaths.

The contrast with India and sub-Saharan Africa is actually quite remarkable. These regions display an excess of *female deaths*, that excess reaching a peak in near-mirror sychronization to the 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. By then, as we shall argue in our computation for missing women, the damage is already done, though from very different sources across China and India.

[28]Death rates may be open to the opposite criticism that they represent current trends rather than an amalgam of past history. (However, they do get rid of the emigration problem.) But the current situation may be a better harbinger of things to come.

Age	Developed Countries			India			China			sub-Saharan Africa		
	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**

TABLE 3. DEATH RATES BY AGE, 2000. "Male" = male death rates, "Female" = female death rates, both per 1000, "Relative" = Male/Female. Source: *World Health Organization and U.N. Population Division.*

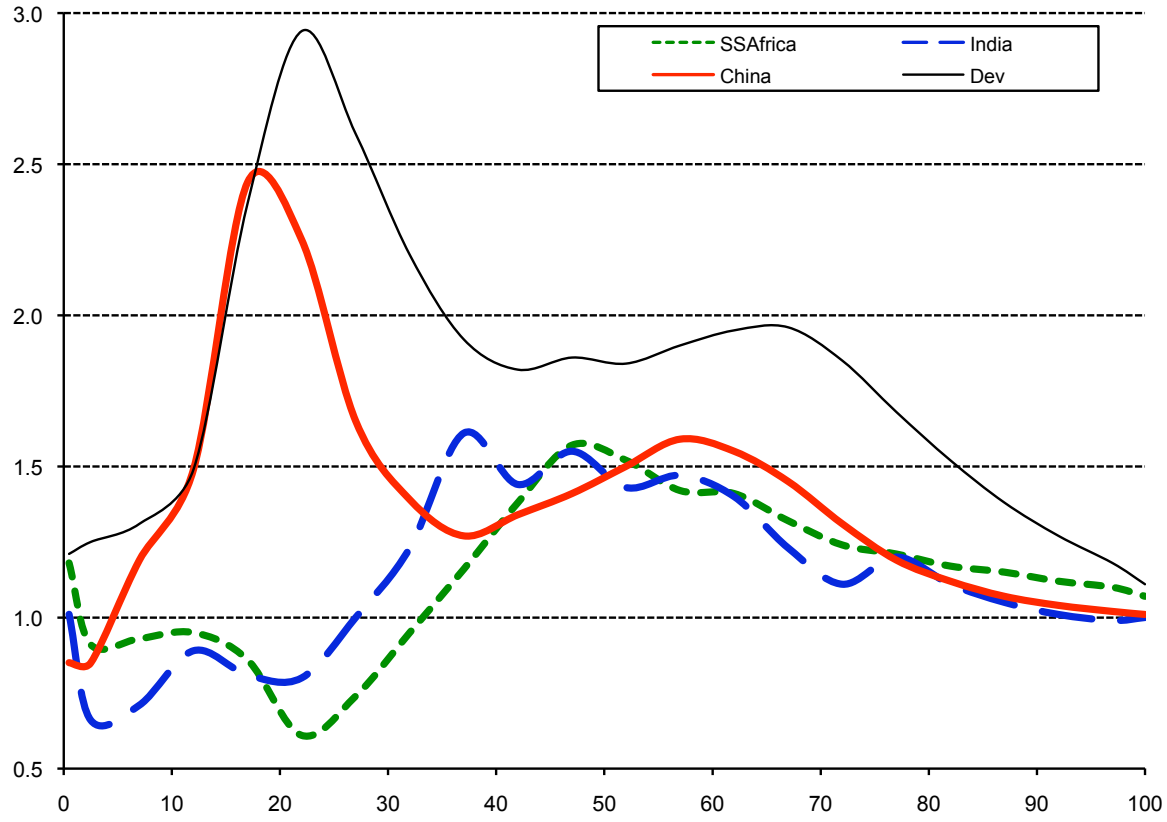

FIGURE 2. MALE-FEMALE RELATIVE DEATH RATIOS BY AGE

We remark again on alternative data. The relatively low male-to-female mortality ratio for younger ages in India is echoed in the recent National Family Life Survey (2005–2006). For the most recent cohort of children (born between 2000 and 2006) the ratio for ages less than 1 is 1.052, whereas for children aged 1 to 4, it is 0.782.[29] Garenne (2003) examines child mortality rates in Sub-Saharan Africa using 60 DHS surveys from that region.[30] He finds that the average male to female mortality ratio for age group 0 to 4 to be 1.10. In Garenne's estimates, the intial excess male mortality significantly declines between the neonatal period (within 28 days after birth) and the post natal period (29 days to 11 months after birth), and then again in early childhood. It is this pattern of an increase in excess female mortality at the youngest ages for sub-Saharan Africa which is consistent with the estimates of 3, and like India, stands in stark contrast to the developed countries.

The U.N. Demographic Yearbooks provide recent information on death rates by age and sex for appoximately ten countries in sub-Saharan Africa. The relatively low male to female

[29]For the cohort born between 1995 and 2000 the corresponding numbers are 1.126 and 0.726. It must be pointed out though that these estimates are formed from only 1500 to 1700 deaths for each sex in the age group 0–1 and 300 to 600 for the age group 1–4. The observations become even fewer if we try to compute mortality rates for children older than 5.

[30]Nine of these are World Fertility Surveys.

mortality ratio for ages 15 to 29 in 3 for Sub-Saharan Africa is also confirmed for the majority of these countries.[31]

3.3. **Missing Women: A First Pass.** We now put together the previous discussion to form an estimate of missing women by age. Recall our accounting scheme; the number of missing women at age $a \geq 1$ is given by

$$\text{mw}(a) = [d^w(a) - u^w(a)]\,\pi^w(a) \tag{8}$$

where $\pi^w(a)$ is the starting population of women of age a, and $u^w(a)$ is the "unbiased" death rate for women at age a. When $a = 0$,

$$\text{mw}(0) = \left[\frac{\sigma(0)}{\widehat{\sigma(0)}} - 1\right]\pi^w(0). \tag{9}$$

Add these all up to generate a first estimate for missing women:

$$\text{mw}_A = \sum_{a=0}^{n} \text{mw}(a). \tag{10}$$

Table 3 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 2.4. For India, recall that estimated sex ratios at birth lie between 1.070 and 1.078, with a DHS average of 1.074 over the years 1996–2006. If we use the developed-country median as a reference sex ratio of 1.059, we assess the number of missing females at birth in India to be equal to approximately 184,000 for the year 2000.[32]

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 almost surely on the low side. We feel reasonably comfortable with the 184,000 estimate using DHS averages as well as 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 a number of missing females at birth in China equal to 885,000 for the year 2000.[33] This may be on the high side if the Chinese census systematically underreports female births. The lowest conceivable number is obtained by using the lower U.N. 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 U.N. 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 U.N. estimate is for 1989, but we will go with it.

---

[31] These include Namibia, Bostwana, South Africa, Swaziland, and Zimbabwe.

[32] Naturally, this requires us to estimate $\pi^w(0)$. We use the UN estimated birth rate for India in the year 2000 of 25.8 births per 1000 population. 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.

[33] 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.

Excess Female Deaths, 000s	India	China	ssAfrica
At Birth	184	644	0
0–1	146	109	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

TABLE 4. EXCESS FEMALE DEATHS BY AGE, 2000. Sources: United Nations and World Health Organization, as well as Table 1. Numbers do not sum to total because of rounding error.

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 1.030–1.035. 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 4 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. Clearly, this runs contrary to the

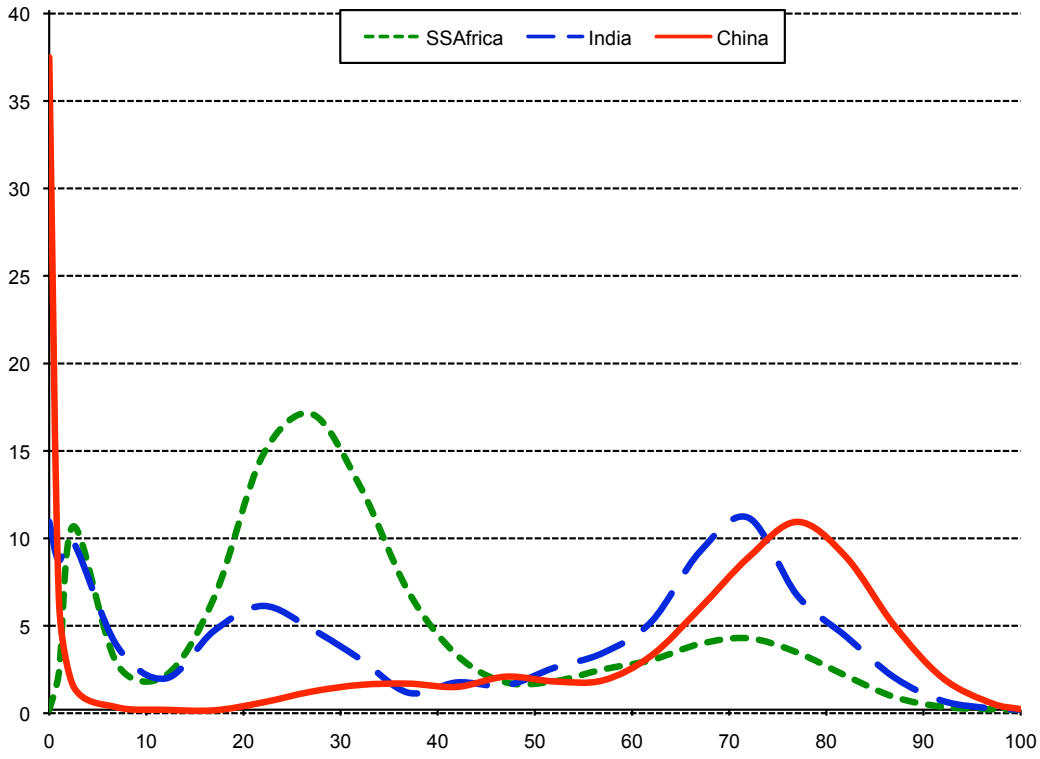

FIGURE 3. MISSING WOMEN DISTRIBUTED BY AGE (IN %)

observations made in much of the earlier literature to the effect that the overall sex ratio in sub-Saharan Africa is low because males and females are treated more equally there.

However, we are not the first to make the assertion that there are missing women in sub-Saharan Africa. Klasen (1996) also points out that the sex ratio at birth is low in sub-Saharan Africa, and uses an appropriately corrected comparison ratio. Apart from this change, he sticks to Coale's method of using West model life tables to construct a counterfactual (see also Klasen and Wink (2003, Table 3)). The numbers, however, are far lower (relative to India and China) than what we obtain here.[34]

It is important to note that we uncover these results, and to this degree, precisely because we decompose the problem by age groups. (We will get an even clearer picture when we decompose by age and disease.) This point is reinforced by the fact that we impute no missing females at birth to sub-Saharan Africa. The female deficit occurs at later ages.

We also recognize that we could have chosen the plausibly higher numbers of missing females at birth for India and China, which are 233,000 and 885,000 respectively. If we

[34]However, Klasen and Wink estimate stocks while we estimate flows. As the discussion in Section 3.4 below suggests, the Chinese *stock* of missing women is likely to be much higher than that for India and sub-Saharan Africa.

modify the totals accordingly, the percentage of missing females (relative to the female population) works out to 0.35 and 0.31 respectively, still way below the corresponding number for sub-Saharan Africa, which is 0.47.[35]

Next, Table 4 makes the observation that while India and China are quite similar in the overall numbers and percentages of missing women, their distribution across age in the two countries are quite different. Figure 3 summarizes this observation by plotting the percentages of missing women that can be "attributed" to different age groups in the three regions.

It is evident that China exhibits a huge spike for missing females at birth: 37% of all missing women in China can be attributed to pre-natal 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. (The figure uses the numbers reported in Table 4.) In contrast, under 11% of the missing women in India are pre-natal (and that number would be at most 13% if we used the higher estimate). Indeed, 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 early 20s (the early 30s, if one uses the larger Chinese estimates). Indeed, 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, of course, over 80%.[36]

Now, this is not to suggest that the emphasis on factors such as infanticide and sex-selective abortion in India is unwarranted. It is well known that there are regions in India where such factors play a very important role. The point is that in India, there has been a *comparative* 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 pre-natal factors, infancy and early childhood.

3.4. **A Remark on Converting Flows to Stocks.** In closing this section, we ask ourselves what it would take to convert these flow numbers to stocks. We would then have a number that may be compared to the Sen-Coale estimates. To this end, consider the proportion of missing women at any age a, which is proxied by the difference between the actual female death rate and its unbiased version; call this $\Delta(a)$:

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

In computing the *flow* of missing women at age a, we simply multiply $\Delta(a)$ by the female population at age a, as we've already done. However, the impact of $\Delta(a)$ on the overall *stock* of missing women is more far-reaching: every cohort of women currently older than a was "affected" by $\Delta(a)$ as they passed through age a sometime in the past. To be more precise, they were affected by *their* version of $\Delta(a)$, the relevant proportion when they were of age a.

[35]Age-specific mortality rates by gender available for a few Sub-Saharan African countries from recent U.N. 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 to 0.63 for those countries.

[36]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.

But we lack data on their version. By applying the same value of $\Delta(a)$ to these older women, we are making the assumption that all age-specific death rates have remained constant over several years. We are extremely reluctant to make this assumption, and therefore postpone this calculation to future work.

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 20m missing women in India, 58m missing women in China and 8m 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.

4. Missing Women and the Epidemiological Transition

We've 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. As we shall now discuss, the changing nature of the disease environment could, in principle, account for much of this. The *epidemiological transition* (Omran (1971)) refers to a transformation in the cause composition of mortality by which acute infectious and deficiency diseases are gradually replaced by chronic and degenerative diseases as the leading causes of death.[37]

Historically, improvements in public health (from the mid-nineteenth century onwards) began in Western Europe and the United States with the major breakthroughs of the germ theory of disease, and with the discovery of penicillin and other vaccines. This led to a substantial decline in mortality. Up to around 1940 most of this progress was confined to rich countries, but the benefits began to percolate thereafter to all countries. The resulting demographic transition is, of course, well-known. What we want to focus on here is the associated *compositional* change in the nature of disease.

The World Health Organization divides the causes of death into three broad categories: (1) communicable, maternal, perinatal, and nutritional diseases; (2) noncommunicable diseases; and (3) injuries. In many developing countries, infectious disease continues to be a major killer. More generally, infectious disease as well as nutritional and reproductive health problems — the group 1 diseases — predominate in higher mortality populations. These are replaced by chronic and degenerative diseases in low mortality populations. By and large, group 1 diseases account for a relatively small proportion of deaths in industrialized countries. Chronic noncommunicable disease in group 2, such as heart ailments or cancer, takes over. 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).

[37] The implicit linearity in this definition may be too simplistic: several infectious diseases, such as tuberculosis, can certainly make a reappearance at "advanced" stages of economic development. Moreover, entirely new infectious diseases, such as AIDS, can also appear.

A possible connection with relative mortality rates for males and females might arise in the following way. Suppose that infectious diseases do not discriminate between males and females in the same way that communicable diseases do. Then gender differences in mortality must widen as countries pass through the epidemiological transition. Indeed, there would be a positive relationship between development and relative male mortality ratios at different ages (Preston (1976), Nathanson (1984)). For this reason, it is not at all surprising to posit an *a priori* link between overall sex ratios and the epidemiological transition. Other compositional effects would presumably include the disappearance of maternal mortality and a possible rise in excess male mortality from risky activities (such as driving motor vehicles).

On the other hand, the observed positive link between relative male mortality and development could arise for reasons that have little or nothing to do with disease composition. It may well be that relative male mortality climbs (with development), and it does so disaese by disease. Below, we attempt to entangle which of these effects is the stronger.

4.1. **Epidemiological Transition: The Changing Face of Death.** Table 5 describes the distribution of death by disease (for the year 2000) in our regions of interest. There is clear evidence for the various phases that describe the epidemiological transition. Sub-Saharan Africa exhibits all the pre-transition characteristics, with Group 1 diseases heading the list (bolstered by HIV/AIDS). India, China and the group of developing countries we consider furnish more snapshots of the transition on the cross-section, with China significantly closer to the developed regions than it is to either India or sub Saharan Africa. To be sure, the process is far from entirely linear. For instance, it is true that in many ways, China has almost completed its epidemiological transition. The pre-revolutionary health situation was very poor, and there have been remarkable improvements since then. By the 1990s, life expectancy had doubled. Indeed, during the Deng era 1960–1980, China moved through the late stages of the transition utilising advances in public health care and medicine (Cook and Dummer (2004)). Yet it is possible to argue that China's rapid economic growth in the last two decades has come at a heavy environmental price, with serious implications for the incidence of respiratory disease. This is apparent in the table above where close to 20% of deaths in China may be attributed to respiratory problems. It is a safe prediction that India will exhibit no different a pattern as it continues to urbanize.[38]

As already discussed, the epidemiological transition merits consideration, for it is entirely possible that a divergence in sex ratios between developing and developed countries may simply be due to changing compositions of disease. Such excess deaths arise not from lack of "similar care" for men and women, but simply from the changing nature of the disease environment. We would then be back to the standard hypothesis that most missing women

[38]In a similar vein, cardiovascular disease is quickly becoming a burden in *developing* countries even before they have rid themselves of infectious diseases (Reddy and Yusuf (1998)). This is evident from Table 5 in the case of India, where heart disease far outstrips cancer as a cause of death compared to the developed world. Improvements in nutrition and health status as well as the successful eradication of major killer diseases have contributed to the ongoing epidemiological transition in India (Gupte et. al. (2001)). Nevertheless, as observed by the World Health Organization, the burden of disease as measured by "premature death" there is second only to sub-Saharan Africa.

Disease	sSAfrica	India	China	Dev
1. Communicable, Maternal, Perinatal, Nutritional	**72**	**40**	**12**	**7**
A. Infectious and parasitic	53	20	5	2
1. Tuberculosis	*3*	*4*	*3*	*0*
2. HIV/AIDS	*20*	*3*	*0*	*0*
3. Diarrhoeal	*7*	*4*	*1*	*0*
4. Childhood Clusters	*7*	*2*	*0*	*0*
5. Malaria	*10*	*0*	*0*	*0*
B. Respiratory Infections	10	10	3	5
C. Maternal	2	1	0	0
D. Perinatal	5	7	3	0
2. Noncommunicable	**21**	**50**	**77**	**87**
A. Malignant neoplasms	4	7	19	26
B. Cardiovascular	10	27	33	38
C. Respiratory	2	6	16	6
D. Digestive	2	3	4	4
3. Injuries	**7**	**10**	**11**	**6**

TABLE 5. DEATHS BY DISEASE, 2000 (%). Source: *Global Burden of Disease (2002).*

are to be found at the pre-natal stage or in early childhood. That is why it is important to examine the transition.

4.2. **Missing Women by Disease.** In what follows, we use data from the Global Burden of Disease Study (GBD), initiated in 1992, which is a major collaborative effort between the Harvard School of Public Health, the World Health Organization, and the World Bank.[39] It is important to describe the data and its limitations, before turning to our computations.

4.2.1. *Data.* 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.[40] The estimates reflect all of the information currently

[39]Refer to Mathers *et. al.* (2004) for details.

[40]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.

available to the WHO.[41] 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.

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 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 underregistration, and adjusted with information from WHO technical programs 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 programs 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 deaths 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 programs. 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.[42] 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

[41]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.

[42]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 a medical personnel, and verbal autopsy reports. This was the case for the Ivory Coast, Ghana, Kenya, Madagascar, and Senegal.

of HIV and AIDS mortality have been developed by UNAIDS 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.[43] 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. 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 (no one 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 assesments 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.

We work with the 2002 GBD study, which provides data for the year 2000. Our objective is to assess missing women by disease, as well provide an estimate of the number of missing women attributable to the changing composition of disease. The importance of this question cannot be overstated. Indeed, if disease composition does play a significant role, it becomes difficult to describe such women as missing: the "similar care" counterfactual does not apply, and explanations based on "discrimination" cannot be easily invoked in those circumstances. Conversely, if most excess female deaths are explicable on a disease-by-disease basis (and not a change in the disease mix), then it becomes that much harder to reject the discrimination viewpoint.

4.2.2. *Methodological Considerations.* Section 2 already describes the procedure for a suitable computation. Suppose that we have data on death rates by age and disease for males and females: $d^m(a,k)$ and $d^w(a,k)$. We compute an "unbiased" death rate for women by age and disease; recall (3):

$$u^w(a,k) = \frac{d^m(a,k)}{\widehat{d^m}(a,k)/\widehat{d^w}(a,k)}. \tag{11}$$

where the hats denote corresponding rates for the reference region. Then excess female deaths by age and disease are given by

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

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 (refer to

[43]Deaths due to landmines and unexploded ordinance were estimated separately.

Section 2.5 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 procedure. The most significant of these is maternal mortality, for which a male death rate is not defined. There is, therefore, no way in which we can use (11) to arrive at an unbiased rate of death from maternal causes. We therefore simply proxy the unbiased 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 multiplying this ratio by the female mortality rate in the same age group for the country in question. That is,

$$u^w(a,\mathrm{mm}) = \frac{\widehat{d^w}(a,\mathrm{mm})}{\widehat{d^w}(a)} d^w(a), \tag{12}$$

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

A second set of exceptions have to do with diseases for which relative death rates in the GBD study for developed country by age are statistically 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.[44] For malaria, the total number of deaths over all ages and over *all* developed regions was less than 100 in the year 2000. It would be absurd to establish reference death ratios in this case; small changes in these numbers can obviously cause relatively large swings in our estimates of missing women. There are similarly too few deaths from childhood cluster diseases in developed countries.

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, particularily in the case of diarrhoeal deaths. We therefore cannot form reliable reference ratios from developed regions in the younger age categories in this case. To address these issues, we use instead the overall death rates from all infectious diseases (excluding HIV/AIDS and other STDs) within each age group in developed regions to compute our reference death rates.[45]

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

[44] In our computation of missing women by age and disease we consider all such categories for which there are at least 2000 female deaths in our country of interest.

[45] 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.

is 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.[46]

Ideally, one would like to to use medical data to arrive at reference rates for mortality by gender for those diseases in which developed-country data is sparse. However, there is an implicit circularity here: to trust the medical estimates, which are often obtained, of necessity, 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.

4.2.3. *General Observations.* We report our estimates in Tables 6 (India), 7 (sub-Saharan Africa) and 8 (China). We begin with some general observations that will assist in reading and interpreting these tables.

In each table, the three main headings are, as before, Group 1 and 2 diseases, and Injuries. The numbers in bold, along the corresponding rows, report missing women for various age categories. In an "ideal" 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 all individual 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* ailment, and then calculating all missing females for that composite ailment. The same is true of Group 2 diseases and Injuries. However, the total, reported as mw_B on the penultimate line of each table, *is* obtained by 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 are in effect freezing the disease composition to that in the country of interest, and therefore picking up, as it were, the "within-disease" component of missing women.

Of course, the fact that we treat each disease-group as a composite ailment means — by the same token — that we are unable to separate out the effect of changing compositions within each subgroup. However, the following observations are relevant. First, the epidemiological transition makes no particular prediction about (nor indeed, any particular reference to) the changing composition of disease within each subgroup. Second, despite the paucity and general inaccuracy of the data, the numbers obtained by summing across the diseases in each subcategory 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.[47]

4.2.4. *India.* With these remarks, we turn to a description of our findings, starting with the estimates for India. The very last line of Table 6 recalls our earlier estimates for missing women by age, with no explicit accounting of disease (mw_A). The penultimate line records mw_B as discussed above, providing estimates for every age category as well as a total,

[46] 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.

[47] In particular, the estimate for "Injuries" matches perfectly with the summation across the two subcategories of "intentional" and "unintentional" injuries, at least when rounded to the nearest thousand.

Excess Deaths, 000s	**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 = **1,637**	**320**	**64**	**191**	**100**	**139**	**236**	**318**	**83**
mw_A = **1,712**	**310**	**93**	**258**	**93**	**120**	**241**	**300**	**113**

TABLE 6. MISSING WOMEN BY AGE AND DISEASE; INDIA, 2000. Source: *Global Burden of Disease* (2002). *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 totaled 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 4.

obtained by adding in the estimate for missing girls at birth. The correspondence between the two rows is remarkable, because (as we've already explained) 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. Invoking Observation 1, we must conclude that practically none of the missing women in India arise from the changing composition of disease (relative to developed regions). It is hard to dismiss the proposition that most of the missing women phenomenon is indeed due to lack of "similar care".

Several other observations are of interest. First, while there are missing women in all age/disease categories, it is evident that the bulk of those missing in the younger age categories come from Group 1 diseases. Group 1 disease between ages 0 and 4 accounts for fully 260,000 missing females, which is over 15% of the total. Of these, about half is accounted for by infectious disease, while the remainder may be attributed to respiratory and perinatal ailments. This is an enormous discrepancy that young girls suffer. To provide an idea of how important this number is, compare the excess deaths from maternal mortality in India, which is widely acknowledged to be a serious issue (see, e.g., Ronsman and Graham (2006)). Maternal deaths account for about 130,000 excess female deaths, no small number to be sure but of the same order of magnitude as *excess* female deaths caused by infectious and parasitic diseases within the age 0–4 category *alone*.

As we've already seen, much of the Indian discrepancy is to be found at older ages. That excess is certainly reflected in Table 6, but the table reveals that at this age range the excess burden falls mainly on noncommunicable Group 2 diseases. Cardiovascular disease represents a particularly important instance. Women die at a rate closer to men from cardiovascular disease relative to developed countries.[48] For instance, 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. But it's not just that: the same excess is found again 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), Drèze and Chen (1992), and Kochar (1999)). Table 6 fully supports this attention by computing excess female mortality from specific disease groups in these age categories.

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 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.

A 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

[48]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.

for 38,000 excess female deaths; the latter for over 13,000; again, these are large numbers, comparable to excess deaths from injuries.

4.2.5. *Sub-Saharan Africa.* Table 7 reports cor responding 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 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.

Where do the excess female deaths in sub-Saharan Africa lie? It appears that the majority of such deaths 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). Like 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 — and sub-saharan Africa stands out from the other two regions in this regard — 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 "similar care" benchmark used to measure missing women. Are these 600,000 excess deaths due to "discrimination" or to different pathways of transmission, perhaps stemming from diverse cultural norms regarding sexuality? For instance, it is well known that heterosexual transmission of HIV/AIDS is common in sub-Saharan Africa, in contrast to modes of transmission in several developed countries, and the resulting discrepancy in relative death rates across gender may account for these excess female deaths. Or coercion may play a large role, or differential treatment. At this stage, we don't know. What we do is draw attention to this channel. Here is a category — one that accounts for a horrific number of deaths — which impacts differentially on women. Ours is a first estimate of just how large that differential might be.

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 play the dominant role. The cardiovascular disadvantage for women (relative to their counterparts for developed countries) makes its presence felt, just as it does in India, though the relative contribution

Excess Deaths, 000s	**0-4**	**5-14**	**15-29**	**30-44**	**45-59**	**60-69**	**70-79**	**80+**
1. 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
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	–	–	–	–	–	–	–
3. 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**

TABLE 7. MISSING WOMEN BY AGE AND DISEASE; SUB-SAHARAN AFRICA, 2000. Source: *Global Burden of Disease* (2002). *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 totaled less than 2000 in this category. mw_B calculated by adding the numbers for Groups 1, 2 and 3 by age.

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

Excess Deaths, 000s	**0-4**	**5-14**	**15-29**	**30-44**	**45-59**	**60-69**	**70-79**	**80+**
1. 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	–	–	–	–	–	–	–	–
2. 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	–	–	–	–	–
3. 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 = **1,592**	**158**	**7**	**10**	**57**	**65**	**111**	**311**	**223**
mw_A = **1,727**	**132**	**2**	**24**	**73**	**89**	**154**	**336**	**272**

TABLE 8. MISSING WOMEN BY AGE AND DISEASE; CHINA, 2000. Source: *Global Burden of Disease* (2002). *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 totaled 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 4.

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".

4.2.6. *China.* Table 8 reports analogous estimates for China. As we've 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 men from these diseases relative to developed countries. We reiterate that it is not just a question of which diseases are the biggest killers.

Overall, the percentage of missing women due to noncommunicable 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, which is much further along in its epidemiological transition.

Finally, as far as "Injuries" are concerned, the situation seems to lie somewhat further away from sub-Saharan Africa (where there are no excess female deaths) and closer to India (where "Injuries" form a large component of such 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.

Does China also uphold our contention that the composition of diseases has little to do with excess female mortality? We believe it does. In the two major age groups that account for (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: a changing composition of disease appears to have reduced rather than exacerbated excess female mortality. There is a definite effect of disease composition in the 60–69 age category, where disease-by-disease comparisons account for a bit over 2/3 of the missing women. Otherwise, in the older age categories, the epidemiological transition does not appear to explain all that many missing females, in line with our observations for India and sub-Saharan Africa. Certainly, if we go by the overall numbers, the transition explains under 8% of all missing women in China.

4.3. **More on Missing Women and the Epidemiological Transition.** The preceding computations indicate that the epidemiological transition does not explain missing females in developing countries. We conclude this section by looking at more aggregate information, which is also consistent with the hypothesis that changing disease composition plays a minor role.

Consider a simple breakdown into Group 1 and Group 2 diseases. Table 9 lists death rates (and relative death rates) by gender under these two headings for the regions of interest: developed countries, China, India and sub-Saharan Africa, as well as relative death rates.

Region	Early Childhood, 0–4			All Ages		
	Male	Female	M/F	Male	Female	M/F
Developed, Gr. 1	0.69	0.55	**1.25**	0.60	0.62	**0.97**
China, Gr. 1	4.59	6.56	**0.70**	0.85	0.79	**1.07**
India, Gr. 1	16.77	17.99	**0.93**	4.01	3.68	**1.09**
sub-S Africa, Gr. 1	39.44	36.47	**1.08**	11.88	10.93	**1.09**
Developed, Gr. 2	0.50	0.44	**1.30**	7.61	7.46	**1.02**
China, Gr. 2	1.44	1.65	**0.87**	5.41	5.42	**1.00**
India, Gr. 2	1.36	1.85	**0.73**	4.93	4.80	**1.03**
sub-S Africa, Gr. 2	1.12	0.95	**1.19**	3.39	3.46	**0.98**

TABLE 9. DEATH RATES PER THOUSAND BY DISEASE GROUP, 2000. Source: *Global Burden of Disease* (2002). Note: Groups 1 and 2 as described in Table 5, except that maternal mortality is excluded.

The definitions follow exactly the description in Table 5, except that maternal mortality is excluded from Group 1.

The figures in boldface are death rates for males relative to females; all other figures in the table are death rates per thousand. Of course, Table 9 agrees with Table 5 in that the death rates for Group 1 diseases increase dramatically as we move from the developed regions, via China and India, to sub-Saharan Africa. (There isn't comparable variation for Group 2 diseases, which is also to be expected.) What is striking, however, are the relative death rates by gender. The last column reports overall mortality rates. The developed countries do, indeed, conform to the view that male-female relative death rates are higher for Group 2 diseases. But *none* of the developing regions follow suit. In fairly sharp contrast, males appear to die relatively *more*, rather than less, within the group of communicable diseases.

A comparison of early childhood mortality (the first of the two boldface colums) is somewhat more in line with the compositional hypothesis. Except for India where the expected rankings continue to be reversed, female death rates are relatively higher for Group 1 diseases.

In any case, this compositional effect across disease groups is comprehensively swamped by the fact that male death rates are so much higher (relative to that for females) in richer countries. Compare, for instance, the early-childhood relative mortality rates (for either group of diseases) in India and China with the same rates for developed regions. Or simply look at the overall figures in the last column for developed regions, and compare these to any of the other figures in that column. *This*, and not a comparison of Group 1 and Group 2 diseases, is the dominant feature of Table 9.

Figures 4 and 5 reinforce this point in the starkest possible way. The figures compare relative death rates for diseases in Groups 1 and 2 across the various regions of interest, for a variety

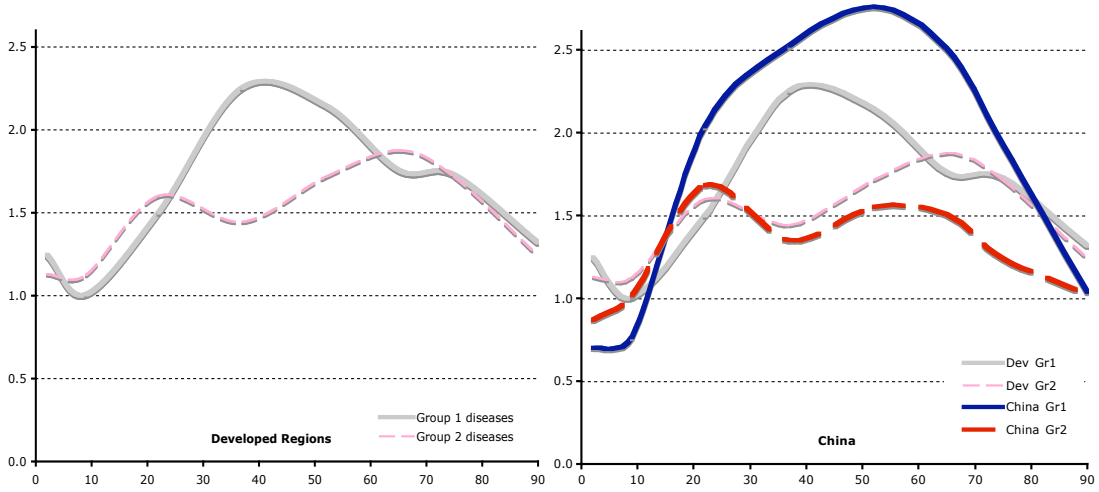

FIGURE 4. MALE-FEMALE DEATH RATIOS BY AGE FOR GROUP 1 AND GROUP 2 DISEASES; DEVELOPED REGIONS AND CHINA. Source: *Global Burden of Disease*. Note: Maternal mortality is excluded from Group 1.

of ages. The first panel of Figure 4 shows relative death rates at different ages for developed regions; in this (as in all panels) the solid line refers to Group 1 diseases, excluding maternal mortality, while the dotted line refers to Group 2 diseases. As already discussed, there is little to choose between the two groups as far as relative mortality is concerned. Moreover, both the lines lie above the 50-50 mark (a ratio of 1.0 in the diagram), displaying the familiar bias in favor of male mortality at all ages.

The second panel, as well as the two panels in Figure 5, conduct exactly the same exercise for China, India, and sub-Saharan Africa. In each of these panels, we've reproduced the graphs for developing regions as pale lines for easy comparison.

At least for India and sub-Saharan Africa, the results speak for themselves. For each disease group, the graph representing relative male mortality rates by age lies almost uniformly below its counterpart for the developed regions. At the same time, in line with Table 9, there is little to choose *across* the graphs for disease groups in any one region. Put another way, these diagrams reinforce our suspicion that the bulk of missing women are due to relative mortality differences *disease by disease*, and not to a change in the composition of disease.

The China panel in Figure 4 is somewhat misleading and requires interpretation. In line with the findings for India and sub-Saharan Africa, the graph for Group 2 diseases lies uniformly below its counterpart for developed regions. But the corresponding graph for Group 1 does not. It appears that in China, and roughly from adolescence onwards, males die more than females do from Group 1 diseases, relative to developed countries. But this reversal is relatively unimportant: in early childhood and in old age, where the bulk of Group 1 deaths lie, the relationship is entirely the same as in India and sub-Saharan Africa,

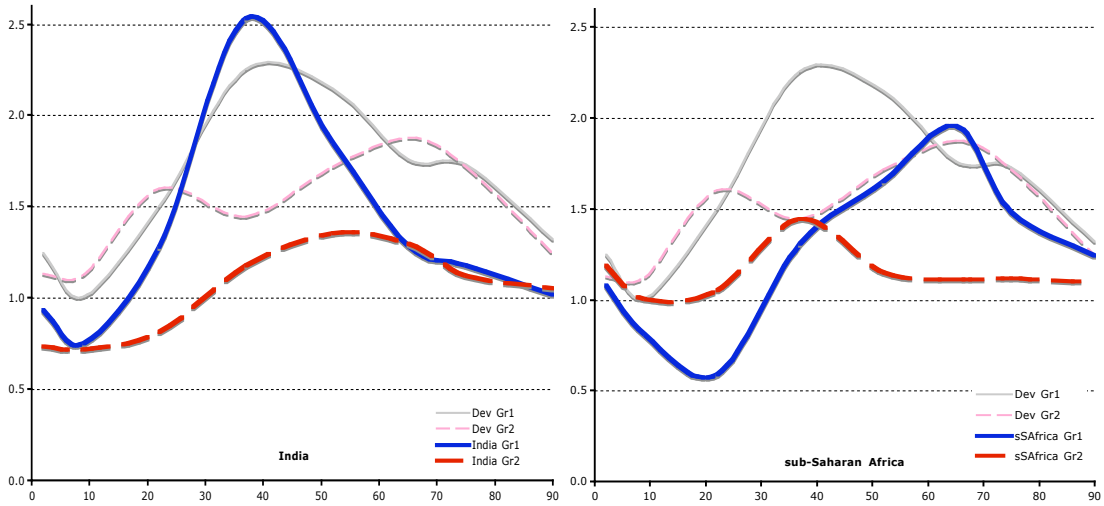

FIGURE 5. MALE-FEMALE DEATH RATIOS BY AGE FOR GROUP 1 AND GROUP 2 DISEASES; INDIA AND SUB-SAHARAN AFRICA. Source: *Global Burden of Disease*. Note: Maternal mortality is excluded from Group 1.

and indeed, Table 9 confirms that the *overall* average of relative death rates from Group 1 for China is almost the same as for India and sub-Saharan Africa.

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, thanks in large part to the availability of Historical Vital Statistics for that time period.

Recall that the sub-Saharan and Indian lines for sex ratios are very flat with age. It turns out that the sex-ratio/age profile for the United States in 1900 was even flatter. We do not report these figures as they are likely to be contaminated by immigration into the United States. But the profile of age-specific relative death rates fully support the contention that males and females died *far* more equally in the 1900 United States than they do now. By exactly the same criterion applied to developing countries today, women were at a relative disadvantage then.

Figure 6 illustrates this quite starkly. The upper curve in that figure plots the male-to-female relative death rates in the developed regions today; it is exactly the same curve as in Figure 2. (Indeed, a comparison with that figure may be useful to the reader at this stage.) The lower curve depicts the same profile for the United States in 1900. The figure speaks for itself: as far as relative death rates by age in the United States in 1900 go, the United States then looked much like sub-Saharan Africa and India look today.

The data that go into Figure 6 allow us to quickly form estimates of missing women by age in the United States in 1900. We do so in Table 10, which recalls the three developing

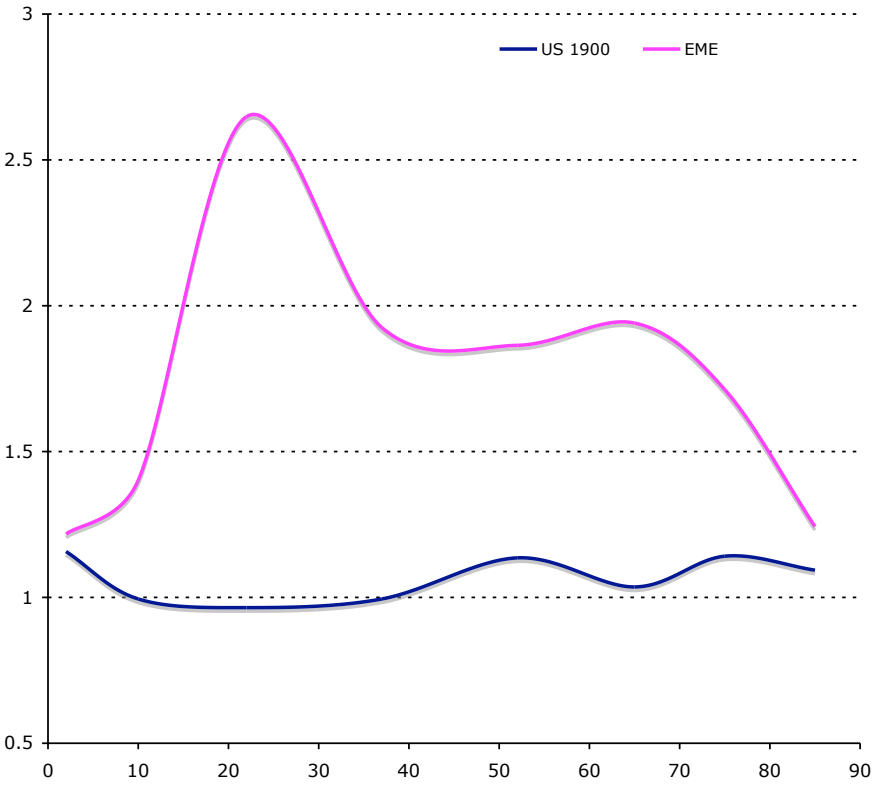

FIGURE 6. MALE-FEMALE DEATH RATIOS BY AGE FOR DEVELOPED REGIONS AND THE UNITED STATES IN 1900. Source: UN, WHO, Historical Vital Statistics of the United States.

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 US was around 37m, while the corresponding number for sub-saharan Africa is around 350m.[49] 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.[50]

Yet our estimates suggest that substantial numbers of women were indeed missing in 1900 United States. As a proportion of the female population the total number of missing women in 1900 in the U.S. 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 the missing women in the United States in 1900 is quite similar to that of India and Sub-Saharan Africa today (simply examine the table with the mental scaling of 10).

There is more to the similarity between 1900 US and sub-Saharan Africa when we look at other health indicators. Table 11 augments Table 5 by including the United States in 1900. The table also adds other indicators such as life expectancy. As is evident from the table, the

[49] A scaling of roughly 15 would put the numbers on par with India, and of around 20 with China.

[50] The overall sex ratio at birth in the U.S. in 1900 is reported to be 1.048; in the year 2000 it is 1.049.

Excess Female Deaths, 000s	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

TABLE 10. EXCESS FEMALE DEATHS IN DEVELOPING REGIONS, 2000 AND IN THE UNITED STATES, 1900. Sources: United Nations and World Health Organization, Table 1 and US Historical Vital Statistics. Note: 1900 US female population approx. 37m.

United States in 1900 fits in rather snugly somewhere between sub-Saharan Africa and India on a number of indicators. It is of particular interest to draw attention to the composition of disease, which reinforces this observation strongly.

The epidemiological transition — in the cross section — leaves a clear footprint through the table. We see that the two leading causes of death in the United States in 1900 were infectious disease (primarily tuberculosis) and respiratory infections. Indeed, we know that death rates from heart disease, tuberculosis, and pneumonia were highly comparable between 1910 and 1917. Then for a short period, 1918 to 1920, pneumonia returned as the main cause of death, after which heart disease became the leading killer, a trend that persists to this day. Between 1900 and 1940 in the United States, the rate of infectious disease and respiratory infection both decreased fourfold, maternal deaths and infancy-related disease both decreased twofold, whereas the death rates from cancer and heart disease both more than doubled.

We are now in a position to attempt to estimate missing women by disease. Table 12 summarizes the results, but some cautionary remarks are in order. 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 (Table 6–8). 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 were completely absent and others like tuberculosis and meningitis were significantly more present. But there are more serious issues of classification: among noncommunicable 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. Thus no

Disease Deaths (%) / Other Indicators	sSA	US 1900	India	China	Dev
Group 1	**72**	**53**	**40**	**12**	**7**
A. Infectious and parasitic	53	33	20	5	2
1. Tuberculosis	*3*	*11*	*4*	*3*	*0*
2. HIV/AIDS	*20*	*0*	*3*	*0*	*0*
3. Diarrhoeal	*7*	*11*	*4*	*1*	*0*
4. Childhood Clusters	*7*	*5*	*2*	*0*	*0*
5. Malaria	*10*	*1*	*0*	*0*	*0*
B. Respiratory Infections	10	15	10	3	5
C. Maternal	2	1	1	0	0
D. Perinatal	5	2	7	3	0
Group 2	**21**	**35**	**50**	**77**	**87**
A. Malignant neoplasms	4	3	7	19	26
B. Cardiovascular	10	11	27	33	38
C. Respiratory	2	1	6	16	6
D. Digestive	2	4	3	4	4
E. Neuropsychiatric	1	7	2	1	5
Injuries	**7**	**6**	**10**	**11**	**6**
Male Life Expectancy (yrs)	48	48	62	71	72
Female Life Expectancy (yrs)	50	51	64	74	79
Overall Death Rate (per 1000)	16	17	7	9	11

TABLE 11. DEATHS BY DISEASE (%) AND MISCELLANEOUS INDICATORS, THREE DEVELOPING REGIONS (2000) AND THE UNITED STATES (1900). Source: Global Burden of Disease (2002) and the Historical Vital Statistics of the United States.

verifiable disease is indeed recorded: death from cardiac fibrillation, a ruptured aneurysm, and perhaps even some perinatal or respiratory conditions were likely all clumped together. We've included apoplexy in the "cardiovascular diseases" in accordance with the present classification system, but be aware that this probably accounts for at least some deaths in other categories.[51] The same can be said for deaths from "convulsions" which has been placed in the category of "neuropsychiatric conditions" under the present classifiation. Given these caveats, bear in mind that Table 12 can only provide rough estimates of the number of missing women by disease.

That said, it is nevertheless of interest that the table unearths a pattern of missing women in the historical data that is not unlike those in developing countries today. The notable exception however is that there appears to be significantly less missing girls (aged 0 to 4) in the historical data. One major reason for this (at least relative to India and China)

[51]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 etiologies. For more on this misclassification, see footnote 55.

Excess Deaths, 00s	**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**

TABLE 12. MISSING WOMEN BY AGE AND DISEASE; UNITED STATES, 1900. Source: Historical Census Reports, National Vital Statistics of the United States. *Notes.* Figures are rounded to the nearest hundred. Group 1 and Group 2 totals are compiled from subcategories (A, B, C, D, not finer subcategories in italics) 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 ratios is computed from an average across all infectious diseases in that age group is used. ** implies that a reference death ratio equal to 1:1 is used. – = no numbers reported because female deaths in 1900 U.S. totaled less than 200 in this category.

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. Another difference is reportedly high death rates from noncommunicable diseases for ages 0 to 4 in 1900 U.S; and there is excess male mortality in

this case. This is quite likely a missclassification 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 are listed according to these terms in today's classification.

One significant drawback of the table is that we are unable to report reliable estimates of missing females due to tuberculosis, which was a major killer in the 1900 United States. The reason is that modern developed regions simply 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. In the table, the reference death ratio for tuberculosis for the younger age groups (less than 30) is simply the average death ratio from all infectious diseases in developed countries in the relevant age group. (Section 4.2.2 discussed these issues in more detail.) If instead we use the 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 TB at the yonger ages in 1900 U.S. 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 for a look at the aggregates. Unlike the developing regions studied earlier, 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 subcategories. The aggregate excess deaths for the two groups are obtained by adding excess female deaths from subcategories A–D (we do not aggregate finer subdivisions as data at those levels are just not comprehensive[52]). 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 12 records these, as also our earlier estimates (mw_A) with the compositional effects included.

In contrast to our findings for 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 2.5 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 subgroups), but possible. So we look for some supplementary evidence for a strong compositional effect.

Figure 7, which is an analogue of Figures 4 and 5, attempts to do this. For easy reference, the figure recalls relative male-female death rates in developed countries for both Groups 1 and 2. The two corresponding plots for the United States in 1900 lie significantly below their developed-country counterparts, just as in the case of modern India and sub-Saharan

[52]The conceptual advantage of building mw_B from the finest subgroupings 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.

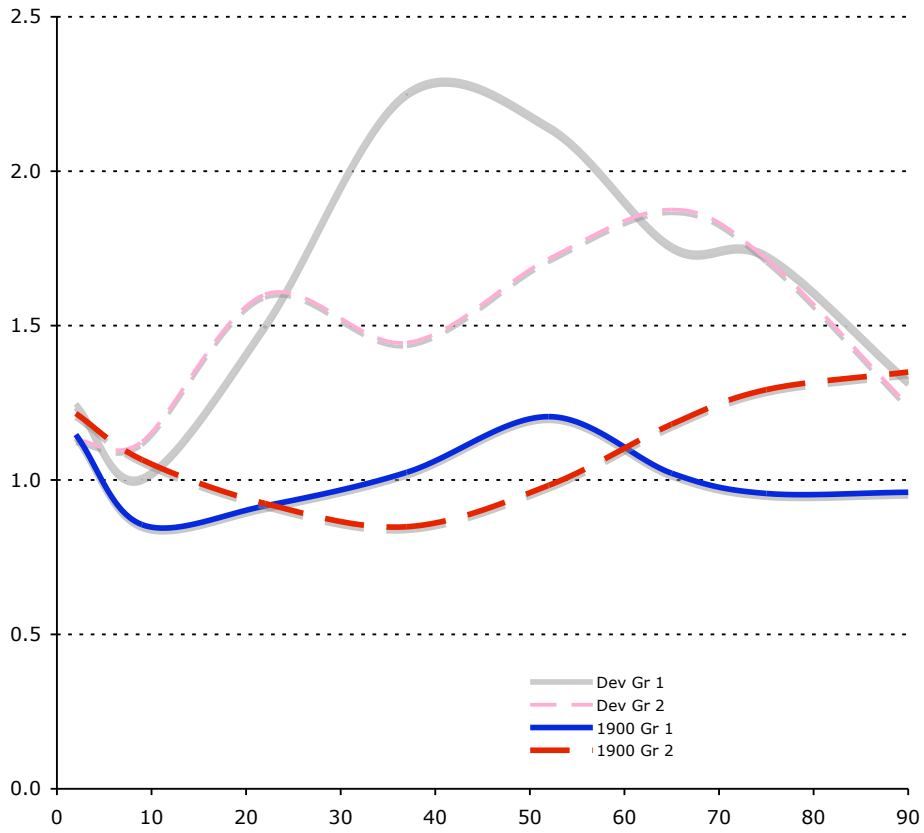

FIGURE 7. MALE-FEMALE DEATH RATIOS BY AGE FOR GROUP 1 AND GROUP 2 DISEASES; THE UNITED STATES, 1900, AND DEVELOPED REGIONS, 2000. Source: *Global Burden of Disease* and Historical Vital Statistics of the United States. Note: Maternal mortality is excluded from Group 1.

Africa.[53] That suggests a large disease-by-disease effect, with females dying relatively quickly at all age groups and for both groups of disease, and this effect is line with what we've seen earlier. What *is* different, however, is that there also appears to be a compositional effect at both low age groups (under 20) and high age groups (above 60) that is eminently in line with the hypothesis suggested by the epidemiological transition. In particular, in *both* these age ranges (which are the age ranges that really matter for death anyway), the male-female relative death rates are lower for Group 1 diseases, compared to Group 2 diseases. A perusal of Figures 4 and 5 reveals that a similar compositional effect is to be found at the very youngest age ranges for China, and at young ages for sub-Saharan Africa (though not

[53]The same is also true of China, except for the middle age segment of Group 1 in which overall death rates are very low anyway.

Region	Early Childhood, 0–4			All Ages		
	Male	Female	M/F	Male	Female	M/F
Developed, Gr. 1	0.69	0.55	**1.25**	0.60	0.62	**0.97**
China, Gr. 1	4.59	6.56	**0.70**	0.85	0.79	**1.07**
India, Gr. 1	16.77	17.99	**0.93**	4.01	3.68	**1.09**
US 1900, Gr. 1	25.39	22.25	**1.14**	7.17	6.84	**1.05**
sub-S Africa, Gr. 1	39.44	36.47	**1.08**	11.88	10.93	**1.09**
Developed, Gr. 2	0.50	0.44	**1.30**	7.61	7.46	**1.02**
China, Gr. 2	1.44	1.65	**0.87**	5.41	5.42	**1.00**
India, Gr. 2	1.36	1.85	**0.73**	4.93	4.80	**1.03**
US 1900, Gr. 2	7.69	6.33	**1.21**	5.13	4.45	**1.15**
sub-S Africa, Gr. 2	1.12	0.95	**1.19**	3.39	3.46	**0.98**

TABLE 13. DEATH RATES PER THOUSAND BY DISEASE GROUP, 2000. Source: *Global Burden of Disease* (2002) and Historical Vital Statistics of the United States. Note: Groups 1 and 2 as described in Table 5, except that maternal mortality is excluded.

anywhere for India), but the historical United States displays this effect at both low and high ages.[54]

To reinforce this point, we also present an extended form of Table 9, now expanded to include the United States in 1900. This is Table 13. Just as we've seen before, the historical United States settles in quite comfortably between India and sub-Saharan Africa, if we use the overall death rate, whether it is for Group 1 or Group 2.[55]

Focus on the last column, which lists relative male-female mortality rates (over all ages) for Group 1 and Group 2 diseases. As discussed earlier (see Section 4.1), each of our developing regions contradicts the idea that males and females are more equally afflicted by Group 1 diseases. But the information we have for the United States in 1900 proclaims it to be quite different: we do indeed see the hypothesized pattern, with death ratios equal to 1.05 under Group 1 and significantly higher — 1.15 — under Group 2. That pattern is also robustly maintained in the early childhood category (the ratios are 1.14 and 1.21 for Groups 1 and 2 respectively). As we have noted earlier, China and sub-Saharan Africa also display this pattern in the 0–4 age category. However, the historical United States is the only one of

[54]Moreover, in the case of the sub-Saharan Africa the bulk of this effect is due to HIV/AIDs. This is *not*, however, a disease that has "disappeared" over the epidemiological transition, so strictly speaking it cannot be attributed to a compositional effect.

[55]An outlier in the overall pattern, however, is the 1900-US rate of child mortality from Group 2 disease, which is way too high relative to what we see today in developing countries. This is almost surely a matter of classification and/or inadequate diagnosis. It turns out that childhood deaths under the "neuropsychiatric" heading account for close to 50% of all deaths in the 0–4 category. This is enormous. The corresponding number for, say, modern India is under 5%. It is clear that several perinatal or respiratory conditions leading to sudden death could easily have been clumped under the neuropsychiatric category.

our four regions of interest that unambiguously exhibits the common wisdom regarding the epidemiological transition, both in childhood and across all age categories. In this single respect, the United States is unlike our three developing regions and is more like developed regions today.

In particular, this last finding is supportive of the suggestion in Table 12 that there is a significant effect of disease composition in accounting for missing women in the historical United States.

6. Summary and Discussion

As we've argued in detail in the Introduction, defining missing women by differences in aggregate sex ratios can be misleading, and even when it is not misleading, it can be uninformative. It is misleading for reasons that are well-known: different countries have different fertility and death rates, and therefore different age distributions. They will also have different disease compositions, and distinct disease groups will, in general, have differential impact on males and females. Countries 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 where the missing women are: at what ages they are clustered, what diseases are responsible. Answering these questions is of profound importance for both theory and policy. Our paper takes a step towards these goals.

We unpack missing women by age, and we unpack further by age *and* disease. While the two procedures are of intrinsic interest, the difference in aggregate numbers between these two procedures can tell us to what extent "lack of similar care", as opposed to simply a change in the disease composition, may be responsible for missing women (Observation 1).

6.1. **A Summary.** Our study of excess female deaths by age yields the following findings:

1. 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.

2. India and sub-Saharan Africa look not dissimilar: while sub-Saharan Africa appears to have no missing females at birth, the corresponding proportion for India is under 11%. More prominently, these regions display missing women at several different ages. Even all the missing females up to age 15 do not account for more than a third of the total in either region.

3. China's missing females, in contrast, are largely pre-natal. About 37–45% of them may be classified as missing at birth. Our study of excess female deaths by age and disease yields the following findings:

4. For developing countries today, the epidemiological transition — the changing *composition* of disease — explains very little of excess female mortality.

5. At young ages, the Group 1 diseases are largely responsible for missing women.

6. The opposite is true at older ages; Group 2 diseases are responsible.

7. In India, respiratory and perinatal illnesses are the main source of excess group 1 female deaths. Maternal mortality is also important but is not of a higher order of magnitude.

8. Cardiovascular deaths are overwhelmingly strong in Group 2 for India, and they dominate all other sources of excess female mortality.

9. 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.

10. In sub-Saharan Africa, *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.)

11. For instance, malaria is an important component of Group 1 disease that accounts for excess female mortality. Maternal mortality is also important.

12. In Group 2, the cardiovascular deficit plays a dominant role.

13. In China, neither Group 1 nor 2 is as important as in India and in sub-Saharan Africa. The dominant source of missing women is pre-natal.

14. In Group 1, the main female killers are respiratory (in the age category 0–4) and perinatal. To us, these are warning signs that active female discrimination in China stretches beyond the pre-natal (compare with point 9 on India).

15. In Group 2, the cardiovascular deficit again plays a dominant role, but respiratory diseases are of comparable importance in excess female deaths, after the age of 60.

16. Indeed, an enormous chunk of missing women are to be found in China, as well as in India, after the age of 60. In China these excess deaths account for close to 40% of all missing women in China. 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.

17. The United States in 1900 appears at once similar and different. It is certainly similar from the pure age perspective: 1900 United States slots right in between sub-Saharan Africa and India, with plenty of missing women (though none at birth). As a fraction of the female population the historical United States has just as many missing women as India or sub-Saharan Africa.

18. However, the United States in 1900 is probably distinct from the disease perspective. Preliminary calculations reveal that the epidemiological transition does account for a significant fraction of missing women. More research is needed to confirm this.

6.2. **The Concept of Missing Women.** Estimates of the number of missing women are meant to represent some measure of the degree of gender discrimination in parts of the developing world. That is, they represent the proportion of women who would not be dead were they to receive nutritional and medical treatment on a similar level to men. This benchmark of "similar care" is complex. In this paper, we've discussed in some detail the question of

age and disease composition, and this inquiry has led to some striking observations and a unified methodological approach. But more work lies ahead 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. 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.[56] This is a case in which our estimates, indirect though they may be, can be extremely useful. Our calculations imply very large numbers in the case of 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. Overall, women are 25% more likely to commit suicide than men and in the rural areas they are 66% more likely. Our indirect estimates may be helpful here as well. Most theories focus on the low status of rural women to explain these startling differences but no concrete evidence is available to date (Phillips et. al. 2002).

Maternal mortality presents another fairly straightforward case. 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. 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. With modernization, perceptions of the process of birth underwent radical changes in the developed world. Birth moved from taking place in intimate communities of women to become another target of science and technology (Rinehart (1987)). Between 1900 and 1930 in the United States, poor obstetric education and delivery practices were mainly responsible for the high numbers of maternal deaths. During the 1930s and 1940s, maternal mortality review committees were established and institutional guidelines for hospital deliveries were developed. Between 1940 and 1950 the number of infants born in hospitals increased from 55% to 90%, and maternal mortality declined by 71%. The legalization of abortion beginning in the 1960s contributed to additional declines (CDC (1999)).

In the developing world today, the leading cause of maternal death is haemorrhage. Whether a woman dies while giving birth depends largely on timely and competent obstetric care. Notwithstanding this, a large proportion of maternal deaths in the developing world occur in hospitals, which demonstrates that improved care in health facilities is still much needed (Ronsmans and Graham (2006)).

[56] Menski (1998) puts the number at roughly 25000 brides who are harmed or killed each year.

Now we take a somewhat more involved example. Consider the large numbers of missing women in sub-Saharan Africa as a result of the HIV/AIDS epidemic. The numbers are nothing short of dramatic: as Table 7 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. Death rates 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.[57] The general and obvious consensus (within a large body of ongoing research) is that in sub-Saharan Africa, the virus propagates largely through heterosexual contact and that women are inherently more susceptible than men to infection from HIV in any given heterosexual encounter. Social norms which allow men a number of concurrent sexual partners will inevitably add to the skew in gender as well as the overall rate of transmission (Bongaarts (1996)).

This example is more complex because it raises the question of what "lack of similar care" means. Is the predominantly heterosexual mode of HIV-transmission in sub-Saharan Africa symptomatic of unequal care? We would conjecture that in a broad cultural sense, the answer is in the affirmative. (For example, the WHO emphasizes the role of social power and violence, so that women are forced to have sex and are often in a situation where they are unable to negotiate for safer sex practices.) But the accounting methodology that we follow is silent on that interpretation; the case has to be made separately.

The cardiovascular deficit presents another sort of interpretative quandary, one that highlights the essentially relativistic nature of our construct. A key finding of our paper 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. The significant gender spread in cardiovascular mortality for industrialized societies did not begin until the 1920s (Nikiforov and Mamaev (1998)). There is an entire array of hypotheses to explain the phenomenon. A class of biological explanations relies on the protective effect of the female sex hormones (or, equivalently, the destructive effect of their male counterparts). Another class of explanations focus on lifestyle differences by gender: diet, attention to personal health and well-being, and so on. Social and economic explanations emphasize the employment of males in more dangerous and stressful jobs. Men apparently engage in more risky behavior (such as cigarette smoking) and such behaviors are then linked to cardiovascular disease.

What, then, constitutes the baseline? Is it really the relatively high male mortality from heart disease in developed countries that is behind the corresponding excess female mortality from this disease in developing countries? Drawing such a conclusion, however, is not at all straightforward. Looking directly at the numbers, the death rates from cardiovascular disease for the age group 70-79 for males and females in India, for example, are 26.17 and 22.20 respectively; the corresponding numbers in developed countries are 15.43 and 9.29. These relative death rates make if very difficult to conclude that women are not at a disadvantage in India. The death rates from cardiovascular disease in India reported here

[57]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.

are consistent with other studies which find that even in rural areas of India, chronic diseases (primarily coronary) are now the leading cause of death (see Joshi *et. al.* 2006). To add to the complication, there is growing evidence that heart disease is more common among South Asians residing in developed countries then in the general population (see Bhopal *et. al.* 1999). Understanding why women seem to be at a particular disadvantage from this disease is important work for future research.

In general, then, 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've taken a preliminary step towards a unified study of a much broader set of issues.

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

References

[1] Abrevaya, J. (2008) "Are There Missing Girls in the United States? Evidence from Birth Data" mimeo, Department of Economics, University of Texas, Austin.

[2] Almond, D. and L. Edlund (2008) "Son-Biased Sex Ratios in the 2000 United States Census," *Proceedings of the National Academy of Sciences*.

[3] Andersson, R. and S. Bergstrom (2007) "Is maternal malnutrition associated with a low sex ratio at birth?" . *Human Biology*, 70(6), 1101-6.

[4] Acemoglu, D. and S. Johnson (2007) "Disease and Development: The effect of life expectancy on Economic Growth" , *Journal of Political Economy*, 115(6), 925-985.

[5] Bhopal, R., N. Unwin, M. White, J. Yallop, L. Walker, K. Alberit, J. Harland, S. Patel, N. Ahmad, C. Turner, B. Watson, D. Kaur, A. Kulkarni, M. Laker, and A. Tavridou (1999) "Heterogeneity of coronary heart disease risk factors in Indian, Pakistanti, Bangladeshi, and European origin populations: cross sectional study". *British Medical Journal*, 319, July 24, 215-220.

[6] Bloch, F. and V. Rao (2002) "Terror as a bargaining instrument: Dowry violence in rural India" , *American Economic Review*, 92 (4), 1029-43.

[7] Bongaarts, John (1996) "Global trends in AIDS mortality" *Population and Development Review*, 22(1), 21-45.

[8] Brass, W. *et al.* (1967) *The Demography of Tropical Africa*, Princeton, NJ: Princeton University Press.

[9] CDC (1999) "CDC on Infant and Maternal Mortality in the United States: 1900-99" *Population and Development Review*, 25(4), 821-826.

[10] Chahnazarian, A. (1988) "Determinants of the sex ratio at birth: review of recent literature" , *Social Biology*, 35(3-4), 214-235.

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

[12] Coale, Ansley (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.

[13] Coale, Ansley, Demeny, Paul, and Barbara Vaughan (1983), *Regional Model Life Tables and Stable Populations*, Princeton: Princeton University Press.

[14] Cook, Ian and Trevor Dummer (2004) "Changing health in China: re-evaluating the epidemiological transition model" *Health Policy*, 67, 329-343.

[15] Cooper, R., B. Osotimehin, J. Kaufman, and T. Forrester (1998) "Disease burden in Sub-Saharan Africa: What should we conclude in the absence of data?", *Lancet*, 351, 208-10.

[16] Das Gupta, M. (2005) "Explaining Asia's "Missing Women" : A new look at the data" , *Population and Development Review*, 31(3), 529–535.

[17] Deaton, Angus (1989) "Looking for Boy–Girl Discrimination in Household Expenditure Data," *World Bank Economic Review* **3**, 183–210.

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

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

[20] Ebenstein, A. (2007) "Fertility choices and sex selection in Asia: Analysis and Policy" , mimeo, Department of Economics, Berkeley.

[21] Gakidou, E., M. Hogan, A.D. Lopez (2004) "Adult mortality: Time for a reappraisal". *International Journal of Epidemiology*, 33, 710-717.

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

[23] Garenne, M. (2003) "Sex differences in health indicators among children in African DHS Surveys". *Journal of Biosocial Science*, 35, 601-614.

[24] Garenne, M. (2004), "Sex Ratios at Birth in Populations of Eastern and Southern Africa," *South African Journal of Demography* **9**, 91–96.

[25] Garg, Ashish, and Jonathan Morduch (1998). "Sibling Rivalry and the Gender Gap: Evidence from Child Health Outcomes in Ghana" *Journal of Population Economics* **11**(4), 471–493. Mimeograph, Department of Economics, Harvard University.

[26] Gupte, M D, V. Ramachandran, and R K Mutatkar (2001) "Epidemiological profile of India: Historical and contemporary perspectives" *Journal of Bioscience*, 26(4), 437-464.

[27] 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.
[28] Hill, K. and J. Trussell (1991) "Further developments in indirect mortality estimation". *Population Studies*, 45, 455-472.
[29] James, W H (1987) "The human sex ratio. Part I A Review of the Literature" *Human Biology*, 59, 721-752.
[30] Jha, P., R. Kumar, P. Vasa, N. Dhingra, D. Thiruchelvam, and R. Moineddin (2006) "Low male-to-female sex ratio of children born in India: national survey of 1.1 million households" , *Lancet*, 367, 211-18.
[31] Joshi, R., M. Cardona, S. Iyengar, A. Sukumar, C Raju, K, Raju, K. Raju, K. Reddy, A. Lopez, and B. Neal (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.
[32] Junhong, C (2001) "Prenatal sex determination and sex-selective abortion in rural central China" *Population and Development Review*, 27 (June), 259-81.
[33] Khoury, M., D. Erickson, and L. James (1984) "Paternal effects on the human sex ratio at birth: Evidence from Interracial Crosses" . *American Journal of Human Genetics*, 36, 1103-1111.
[34] Klasen, Stephan (1996), "ÔÔNutrition, Health, and Mortality in Sub Saharan Africa: Is There a Gender Bias?ÕÕ *Journal of Development Studies* **32**, 913–933.
[35] Klasen, Stephan and Claudia Wink (2003), "'Missing Women': Revisiting the Debate," *Feminist Economics* **9**(2–3), 263–299.
[36] 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.
[37] Lerchl, A. (1998) "Seasonality of sex ratio in Germany" *Human Reproduction*, 13(5), 1401-1402.
[38] Lin, M-J, J-T Liu, and N. Qian (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.
[39] Lin, Ming-Jen and Ming-Ching Luoh (2008) "Can Hepatitis B mothers account for the number of missing women? Evidence from three million newborns in Taiwan", forthcoming in *American Economic Review*.
[40] Mackenzie, C., A. Lockridge, and M. Keith (2005) "Declining sex ratio in a First Nation Community" *Environmental Health Perspectives*, 1131(10), 1295-1298.
[41] Mathers, C., C. Bernard, K. Moesgaard Iburg, M. Inoue, D. Ma Fat, K. Shibuya, C. Stein, N. Tomijima, and H. Xu (2004) "Global Burden of Disease in 2002: data sources, methods, and results", Global Programme on Evidence for Health Policy Dicussion Paper No. 54, World Health Organization.
[42] Mathews, T. and B. Hamilton (2005) "Trend Analysis of the Sex Ratio at Birth in the United States," *National Vital Statistics Reports* **53**, Number 20, June 14.
[43] 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.
[44] Menski, W. (1998) (ed.) *South Asians and the dowry problem*, Vistaar Publications, New Delhi.
[45] Morton, N, C. Chung, and M. Mi (1967) "Genetics of interracial crosses in Hawaii" in *Monograph in Human Genetics* vol. 3, New York.
[46] Myers, R. (1947) "Effect of war on the sex ratio at birth" *American Sociological Review*, 12(1), 40-43.
[47] Nathanson, C. A. (1984) "Sex differences in mortality" *Annual Review of Sociology*, 10, 191-213.
[48] Nikiforov, Sergey. and Valery Mamaev (1998) "The Development of Sex Differences in Cardiovascular Disease Mortality: A Historical Perspective" . *American Journal of Public Health*, 88(9), 1348-1353.
[49] Oster, Emily (2005) "Hepatitis B and the Case of the Missing Women" , *Journal of Political Economy*, 115(6), 1163-1216.
[50] Oster, Emily (2008) "Proximate sources of population sex imbalance in India", forthcoming in *Demography*.
[51] Oster, Emily, Gang Chen, Xinsen Yu, and Wenyao Lin (2008) "Hepatitis B does not explain male-biased sex ratios in China", *mimeo*, Department of Economics, University of Chicago.
[52] Preston, Samuel (1975) "The changing relation between mortality and level of economic development" *Population Studies*, 29(2), 231-248.
[53] Qian, N. (2008) "Missing women and the price of tea in China: The effect of relative female income on sex imbalance" , forthcoming in *Quarterly Journal of Economics*.
[54] Reddy, K S and S. Yusuf (1998) "Emerging epidemic of cardiovascular disease in developing countries" *Circulation*, 97, 596-601.
[55] Rinehart, Sue Tolleson (1987) "Maternal health care policy: Britain and the United States" *Comparative Politics*, 19(2), 193-211.

[56] Ronsmans, C. and W. Graham (2006) "Maternal Mortality: Who, When, Where and Why," *Lancet* **368**, 1189–1200.
[57] Sen, Amartya (1990) "More Than 100 Million Women Are Missing" *The New York Review of Books*, 37(20), December 20, 1990.
[58] Sen, Amartya (1992) "Missing women" *British Medical Journal*, 304 (March), 587-588.
[59] Snow, R., C. Deichmann, K. Marsh (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.
[60] Subramanian, S., and A. Deaton (1991). "Gender Effects in Indian Consumption Patterns," *Sarvekshana* **14**, 1–12.
[61] Sudha, S. and S I Rajan (1999) "Female demographic disadvantage in India 1981-1991: Sex Selective Abortion and Female Infanticide" *Development and Change*, 30, 585-618.
[62] Tarver, J. and C. Lee (1968) "Sex Ratio of Registered Live Births in the United States, 1942–63," *Demography* **5**, 374–381.
[63] Zeng Yi, Tu Ping, Gu Baochang, Xu Yi, Li Bohua, and Li Yongping (1993), "Causes and Implica- tions of the Recent Increase in the Reported Sex Ratio at Birth in China,Ó *Population and Development Review* 19(2), 283–30.