MISSING UNMARRIED WOMEN

SIWAN ANDERSON AND DEBRAJ RAY

ABSTRACT. That unmarried individuals die at a faster rate than married individuals at all ages is well documented. Unmarried women in developing countries face particularly severe vulnerabilities, so that excess mortality faced by the unmarried is more extreme for women in these regions compared to developed countries. We provide systematic estimates of the excess female mortality faced by older unmarried women in developing regions. We place these estimates in the context of the missing women phenomenon. There are approximately 1.5 million missing women between the ages of 30 and 60 years old each year. We find that 35% of these missing women of adult age can be attributed to not being married. These estimates vary by region. India has the largest proportion of missing adult women who are without a husband, followed by the countries in East Africa. By contrast, China has almost no missing unmarried women. We show that 70% of missing unmarried women are of reproductive age and that it is the relatively high mortality rates of these young unmarried women (compared to their married counterparts) that drive this phenomenon.

1. INTRODUCTION

It is a well established fact that in developed countries, married individuals experience lower mortality rates than their unmarried counterparts. It is a relationship that has been studied since the mid 1800s (beginning with the work of William Farr (1858) for France). Zheng and Thomas (2013) emphatically state that "the beneficial effect of marriage on health is one of the most established findings in medical sociology, demography, and social epidemiology.[1] This relative excess mortality for the unmarried occurs at all ages, for both sexes, and for all causes of death (Johnson et. al. 2000, Nagata et. al. 2003). This difference persists, for both sexes, after controlling for observed socioeconomic and health related variables (Randall et. al. 2011). The effect of death of a spouse on the mortality of the survivor — the so-called "widowhood effect" — is well established. The increased probability of death among recently bereaved has been found in men and women of all ages around the world (Subramanian et. al. 2008). None of this should come as a surprise: after all, marriage provides significant economic, psychological and environmental benefits, and it involves two partners caring for each other.

Developing countries are no exception. The data is sparser but the evidence we do have similarly indicates relative excess mortality for the unmarried in most age groups and for both sexes.[2] Arguably, most of this stems from widow(er)hood. After all, in developing countries, marriage at young ages is essentially universal, so that unmarried adults are typically widowed. Moreover, the price of widowhood is particularly steep for women. In South Asia, that marginalization is well documented for both India and Bangladesh (Chen and Dreze 1992; Jensen 2005, Rahman, Foster, and Menken 1992). Increased vulnerability is not only a result of losing the main breadwinner of the household (the husband), but also property ownership laws and employment norms which restrict the access of widows to economic resources.

Date: February 2018.

Vancouver School of Economics and CIFAR. siwan.anderson@ubc.ca.

Department of Economics, New York University, and University of Warwick. debraj.ray@nyu.edu. Ray acknowledges research support under National Science Foundation Grant SES-1261560.

[1]There is correspondingly an extensive literature on the subject: amongst numerous others, see Goldman 1993, Shor et. al. 2012, Robards et. al. 2012, Murphy et. al. 2007, Liu and Johnson 2009, and Rendall et. al. 2011.

[2]See Section 3.2.

Patrilocal norms exacerbate the situation. The economic and social support that a widow receives in her late husband's village is typically extremely limited. Add to these a variety of customs and beliefs: seclusion and confinement from family and community, a permanent change of diet and dress, and discouragement of remarriage. Widows in South Asia are considered to be bad luck and to be avoided; they are unwelcome at social events, ceremonies and rituals. The most infamous (though least widespread) manifestation of these social customs is *sati*, self-immolation on the husband's cremation pyre.

And this is no small matter. In India, there are estimated to be more than 40 million widows, which reflects the large husband-wife age gap (approximately 6 years) and greater remarriage incidence among widowers compared to widows (Jensen 2005).

A similar plight can be documented for African countries (Sossou 2002, Oppong 2006). There too, rules of inheritance and property rights restrict the access of a widow to her late husband's resources.[3] Rituals of seclusion and general isolation of widows are a widespread practice in many parts of Africa. Widows can be accused of witchcraft and persecuted, if suspected to have somehow caused their husbands' death. Witchcraft beliefs are widely held throughout Sub-Saharan Africa and elderly women are the typical targets of witch killings (Miguel 2005). Customarily, causes for any death are sought within the prevailing social system, and suspected witches in the family of the dead or sick are often a prime focus of blame (Oppong 2006).

Given these extreme vulnerabilities faced by widows in developing countries, we expect the excess mortality faced by the unmarried to be *relatively* more extreme for women in these regions. That can lead to excess female mortality among adult women. In this paper, we aim to provide systematic estimates of the extent of such excess female mortality in developing regions.

Our approach and methodology allow us to place our estimates in the context of the "missing women" phenomenon. The concept, developed by Amartya Sen (1990, 1992), is based on the observation that in parts of the developing world, notably India and China, the overall ratio of women to men is inappropriately low. Sen translated these skewed sex ratios into absolute numbers by calculating the number of extra women who would have been alive in a particular country if that country had the same ratio of women to men as in areas of the world with supposedly less gender bias in health. Our earlier work (Anderson and Ray 2010, 2012) examined how missing women were distributed across different age groups, regions, and cause of death. Received wisdom has it that gender bias at birth (say, via sex-selective abortions) and the mistreatment of young girls are dominant explanations. However, while we did not dispute the existence of severe gender bias at young ages, we found that the vast majority of missing women were of adult age.

In this paper, and following on the observations above, we focus on adult female excess mortality between the age of 15 and 69. As already stated, this is a large fraction of overall excess mortality. Indeed, in line with the numbers obtained in our earlier research, our estimates here suggest that there are approximately 3 million missing women of adult age each year. Our objective is to estimate the share of this excess mortality that can be attributed to "non-marriage" *alone*.

We informally describe our methodology. To begin with, observe that unmarried women are relatively more prevalent in developing countries. As one might guess, this is primarily due to higher mortality rates at every age, so that the incidence of widowhood is significantly larger at every age group in developing (relative to developed) countries. As already noted, unmarried women face relatively elevated mortality risk compared to their married counterparts. To be sure, there are widowers as well as widows, and they too are subject to elevated risks of death. Nevertheless, as long as death rates are elevated in this manner, so are the absolute numbers of missing women — provided that the mortality rates are gender-skewed in the region to begin with. Below, we refer to this as the *marriage incidence effect*.

[3]van de Walle (2011) shows how households headed by widows in Mali have significantly lower living standards than other households in rural and urban areas.

Moreover, there are additional distortions caused by variation in the elevation factors themselves. That is, the elevation in female mortality risk stemming from non-marriage could be relatively higher in the region of interest. We call this the *skewed elevation ratio effect*. To take this into account, the entire analysis must turn on a comparison of different ratios, and not just the elevation factors *per se*. We develop a methodology to separate and understand these two components.

It turns out that more than 710,000 adult women who go missing each year can be attributed to non-marriage. These estimates vary by region. India comprises the largest number of missing unmarried women, 230,000 each year. Followed by the countries of East Africa with 150,000. The estimates elsewhere in the world are significantly lower. In China, for example, only 3% of the missing adult women are without a husband. Furthermore, our estimates demonstrate that 70% of the missing unmarried women are of reproductive age (between 15 and 49 years of age). We show that within this socially marginalized group of young unmarried women, the skewed elevation ratio effect is primarily responsible for excess female mortaility. The remaining 30% of the missing unmarried women are older (between 50 and 69). We find that excess female mortality amongst this older unmarried group is due largely to the marriage incidence effect.

The above decompositions are described in detail in what follows. But it is also important to point out what we do *not* do. This is not an exercise that fully identifies the causal channels that conspire to work through marriage — or its absence — in creating excess female mortality. Rather, the exercise is dedicated to showing that the numbers involved are large. Certainly, the vulnerabilities faced by unmarried women in developing countries have been discussed. But a systematic *quantitative* assessment of the problem, especially one that places such vulnerabilities in the larger context of excess female mortality, has not — to our knowledge — been conducted. The contribution, then, is to place a magnitude on this problem in terms of the extreme excess mortality risk that unmarried women appear to experience. That said, in Section 6 we engage in a discussion of the causal impact and the possible bias created by systematic selection effects.

2. Methodology

We first compute the number of missing women at adult ages. We then take into account the role of marital status to generate estimates of excess female mortality among the unmarried.

2.1. **The Basics.** The methodology we employ is in the spirit of the Sen contribution. Any computation of missing women presupposes a counterfactual. For Sen this counterfactual is the set of developed countries and we adopt the same approach here.[4] For each age group we posit a "reference" death rate for females, one that would obtain if the death rate of females in that country were to bear the same ratio to the existing death rate of males as the corresponding ratio for developed countries. We subtract this reference rate from the actual death rate for females, and then multiply by the population of females in that category. This is the definition of "missing women" in the age group under consideration.

More formally, let a stand for an age group. Let $d^m(a)$ and $d^w(a)$ represent the death rates at age a, for men and women respectively, in the region of interest (or "our region"). Use the label $\widehat{}$ for the same variables in the benchmark or reference region.[5] The *reference death rate* for women of age a in our region is defined by the number that equalizes *relative* gender-specific death rates in the region to the corresponding ratio in

[4]We discuss the appropriateness of using developed countries as the benchmark in Anderson and Ray (2010). In that paper, we also considered Latin America and the Caribbean as an alternative benchmark, as we find far fewer missing women in that region of the world. With this alternative benchmark the total number of missing women between the ages of 15 and 69 reduces by about 12%. The World Development Report (2012) used our methodology to compute excess female mortality around the world for every developing country across the different age groups. Their estimates match those of Anderson and Ray (2010) very closely.

[5]We use the group of Developed Countries as defined by the World Bank: Europe, Canada, United States, Australia, New Zealand, and Japan.

the reference region. That is, it is the value $r^w(a)$ that solves the condition $r^w(a)/d^m(a) = \widehat{d}^w(a)/\widehat{d}^m(a)$, or equivalently,

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

This methodology turns a blind eye to the prevailing *level* of the death rates in our region, thereby implicitly acknowledging that the region can be relatively poor and so prone to greater overall mortality. But it demands that whatever that higher death rate might be, the ratio of women dying relative to men should be no different compared to that in the reference region.[6] And if we accept that, then the number of age-specific extra female deaths, or "missing women," in our region in a given year would be equal to the difference between the actual and reference death rates for women, weighted by the number of women in that age group:

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

where $\pi^w(a)$ is the starting population of women of age a. Notice that while the reference death rate is not affected by the average mortality rate, this estimate is: the absolute numbers of missing women would increase with higher average mortality, *ceteris paribus*.

Anderson and Ray (2010) discuss the interpretation of (2) in some detail. In particular, excess female mortality might arise from a number of causes, and only some of these are explicitly interpretable as discrimination against women, the most obvious examples being excess female mortality at the pre-natal stage, or at birth or in infancy. Other factors, such as excess female mortality from cardiovascular disease or HIV/AIDS, need a more nuanced interpretation. We do not revisit these issues here but refer the interested reader to our earlier paper.

2.2. **Marriage Rates and Elevation Ratios.** We describe a strategy for identifying excess female deaths, if any, due to non-marriage. This is a subtle problem. Typically, both unmarried men and women have higher death rates, correcting for age. To the extent that creates larger numbers of deaths all around, it also creates a larger number of *excess* female deaths as well. But there is a second factor at work, which is the *comparative* extent by which the death rates for women are raised. To be sure, these ratios are elevated for both our region as well as the reference region. For our purposes we will need to compare two sets of ratios. All this will be made clearer with a bit more notation and formalism.

There is also the question of finer categories of non-marriage: widowed, divorced, or never-married. Data limitations force us to lump these subcategories together. We postpone further discussion to Section 3.3.

Let $\sigma^m(a)$ and $\sigma^w(a)$ be the incidence of unmarried men and women respectively in age group a. For instance, if $\sigma^m(a) = 0.1$, then 10% of all men in age group a are unmarried; and 90% are married. Denote by $e^m(a)$ and $e^w(a)$ the *elevation factors* for males and females respectively; that is, the relative rise in the death rates conditional on lack of marriage. For instance, if $e^w(a) = 1.1$, then unmarried women in age group a are 10% more likely to die, compared to married women in that age group.

2.3. **Excess Female Mortality by Marital Category.** Our focus is on marriage — or the lack thereof — and to get at this it will be useful to proceed in a couple of steps. Begin by carrying out the same exercise leading up to (1), but starting with *married* individuals. That is, let $\delta^w(a)$ and $\delta^m(a)$ be the death rates for married females and males, respectively. Use the label $\widehat{\ }$ to denote these same variables for the benchmark or reference region. We can now generate a "reference" death rate for married women of age a in our region

[6] We are, of course, aware that any such "adjustment method" can be criticized; see Anderson and Ray (2010) for a discussion of this point. But keeping the ratio constant has a strong intuitive appeal. Besides, this is the approach predominantly taken in the literature since the work of Sen.

of interest by

$$\rho^w(a) = \frac{\delta^m(a)}{\widehat{\delta}^m(a)/\widehat{\delta}^w(a)}, \tag{3}$$

and we are then in a position to define *excess female mortality with marriage benchmarks* (EFM^0) at age a by

$$\text{EFM}^0(a) = [\delta^w(a) - \rho^w(a)]\,\pi^w(a), \tag{4}$$

where $\pi^w(a)$ is, as before, the entire population of females of age a. Note that we are multiplying by the full female population, so this is not an estimate of how many women are missing among the married. It is an estimate of missing women in the entire population *under the presumption that the death rates for married individuals apply to everyone*. So, for instance, if the married and unmarried death rates were all the same for women and men, then EFM(a) = EFM$^0(a)$. But if there is elevation, then EFM$(a) >$ EFM$^0(a)$. We are interested in the empirical magnitude of this difference.

Let us make the connection clearer by converting the values on the right-hand side of (4) into the aggregated rates that we have in (2). Recalling the elevation factors and marriage incidence rates already defined, we see that

$$d^i(a) = \left[\sigma^i(a)e^i(a) + (1 - \sigma^i(a))\right]\delta^i(a) \equiv c^i(a)\delta^i(a), \tag{5}$$

for $i = m, w$, where $c^m(a)$ and $c^w(a)$ can be viewed as "correction factors" that are generated by the elevation factors, and by the proportions of married males and females.

An increase in the elevation factor $e^i(a)$ raises $c^i(a)$. Moreover, given some $e^i(a) > 1$, a greater incidence of non-marriage will also increase $c^i(a)$, as there is a shift away from the lower death rate category. In short, larger values of $c^i(a)$ point to higher death rates in the unmarried category, and given that, is correlated with a lower incidence of marriage.

Use the label $\widehat{\ }$ to denote these same variables for the benchmark or reference region. Invoking (5) for both regions and both genders, we have

$$\begin{aligned}
\rho^w(a) &= \frac{\delta^m(a)}{\widehat{\delta}^m(a)/\widehat{\delta}^w(a)} \\
&= \frac{d^m(a)}{\widehat{d}^m(a)/\widehat{d}^w(a)}\,\frac{1/c^m(a)}{\widehat{c}^w(a)/\widehat{c}^m(a)} \\
&= r^w(a)\frac{1/c^m(a)}{\widehat{c}^w(a)/\widehat{c}^m(a)},
\end{aligned} \tag{6}$$

where $r^w(a)$ is the unbiased death rate for all women in our region at age a, defined earlier in (1). Using (5) and (6) in (4), we must conclude that

$$\begin{aligned}
\text{EFM}^0(a) &= [\delta^w(a) - \rho^w(a)]\,\pi^w(a) \\
&= \left[\frac{d^w(a)}{c^w(a)} - r^w(a)\frac{1/c^m(a)}{\widehat{c}^w(a)/\widehat{c}^m(a)}\right]\pi^w(a) \\
&= \left[d^w(a) - r^w(a)\frac{c^w(a)/c^m(a)}{\widehat{c}^w(a)/\widehat{c}^m(a)}\right]\frac{\pi^w(a)}{c^w(a)} \\
&\equiv [d^w(a) - \theta(a)r^w(a)]\,\frac{\pi^w(a)}{c^w(a)},
\end{aligned} \tag{7}$$

where

$$\theta(a) = \frac{c^w(a)/c^m(a)}{\widehat{c}^w(a)/\widehat{c}^m(a)}$$

can be viewed as the *relative elevation* at age a in the region, compared to the reference region. Recall that $c^i(a)$ is larger the greater the elevation factors and the smaller the incidence of marriage. Moreover, the larger these differences in our region, compared to the reference region, the larger is the value of $\theta(a)$.

The gap between EFM and EFM^0 can be viewed as the additional number of missing women due to non-marriage (and the consequently higher death rates). Call this gap EFM^1, then

$$\begin{aligned} \text{EFM}^1(a) &\equiv \text{EFM}(a) - \text{EFM}^0(a) \\ &= [d^w(a) - r^w(a)]\,\pi^w(a) - [d^w(a) - \theta(a)r^w(a)]\,\frac{\pi^w(a)}{c^w(a)} \\ &= \left[1 - \frac{1}{c^w(a)}\right]\text{EFM}(a) + [\theta(a)-1]r^w(a)\frac{\pi^w(a)}{c^w(a)}. \end{aligned} \tag{8}$$

There are two components in this equation. The first is what one might call the *marriage incidence effect*, and is given by the term $[1-(1/c^w(a))]$EFM(a). Lower marriage — typically because of widow(er)hood — increases death rates for both men and women, and makes $c^w(a)$ larger than 1. This elevation is more accentuated the higher the rate of non-marriage. Even if the relative death rates for women change exactly the same way for both the reference region and the region of interest (so that $\theta(a) = 1$), this will *still* increase the total number of missing women relative to the marriage benchmark. After all, the absolute gap between the death rate and its reference counterpart will have widened.

The second component is what might be termed the *skewed elevation ratio effect*, and is given by the term $[\theta(a)-1]r^w(a)(\pi^w(a)/c^w(a))$ in (8). If female death rates in our region climb with lack of marriage at a rate that exceeds the benchmark rate of the reference region, then this increases the value of $\theta(a)$ and contributes to missing women. (Even though it might increase the denominator $c^w(a)$ as well, the net effect is positive.[7]) We will need to be guided by the data on this matter, but there are reasons to believe that the female correction factor does indeed bear a higher ratio to its male counterpart in the region of interest, relative to the reference region. First, if there is a large gap between male and female age at marriage in the region of interest, adult women will tend to be widowed more often, so that the rate of non-marriage will be higher, especially in the middle-age category. Second, if the traditional discrimination against women is reflected to a proportional degree in widowhood, the relative elevation ratio $e^w(a)/e^m(a)$ will tend to be higher in the region of interest. Both factors work in the same direction: they raise the female correction factor relative to the male factor in the region of interest.

It is also worth reiterating that the correction term c is generated from two sources: one is the ratio of elevation factors, and the other is the incidence of non-marriage. Of course, the latter has no meaning in the absence of the former: if e were 1, then the incidence of non-marriage would be irrelevant. On the other hand, if $e > 1$ (and we shall see that this is indeed the case), then a lower marriage rate raises the correction term. In an effort to disentangle the "pure" effect of the elevation ratio from the additional impact of marriage incidence, we can decompose each component of (8) into two parts. We can do so by first shutting down the differential incidence of marriage altogether, by simply using the rates of marriage that prevail in the reference region. That is, we can define pseudo-correction factors for the region of interest by

$$c_e^i(a) \equiv \widehat{\sigma}^i(a)e^i(a) + (1 - \widehat{\sigma}^i(a))$$

for $i = m, w$, and a pseudo relative elevation, given by

$$\theta_e(a) \equiv \frac{c_e^w(a)/c_e^m(a)}{\widehat{c}^w(a)/\widehat{c}^m(a)}.$$

With these in hand, we can define a new measure of missing women from non-marriage, call it EFM$_e^1$, which allows for the elevation but not the different rates of marriage across the two regions. It is given by

[7] After all, $\frac{\theta(a)-1}{c^w(a)}$ equals $\frac{1/c^m(a)}{\widehat{c}^w(a)/\widehat{c}^m(a)} - \frac{1}{c^w(a)}$.

the following analogue of (8) for each age group a:

$$\text{EFM}_e^1(a) \equiv \left[1 - \frac{1}{c_e^w(a)}\right] \text{EFM}(a) + [\theta_e(a) - 1] r^w(a) \frac{\pi^w(a)}{c_e^w(a)}. \tag{9}$$

It also has two components, of course, just as EFM_e^1 did. The remaining term is

$$\text{EFM}^1(a) - \text{EFM}_e^1(a)$$

as a matter of accounting, and can be tentatively interpreted as the additional number of missing women due to changes in marriage incidence alone. This term may be positive or negative. If non-marriage rates are high in the region of interest, say due to widowhood, this term will also make a positive contribution.

Just how large $\text{EFM}^1(a)$ and $\text{EFM}_e^1(a)$ are is in practice an empirical question, and the goal of this paper is to provide both estimates.

3. Data

In our computations of $\text{EFM}^1(a)$ and $\text{EFM}_e^1(a)$, we focus on the ages 15 to 69. Our regions of interest are India, China, South Asia (excluding India), Southeast Asia, West Asia, East Africa, West Africa, Middle Africa, South Africa, and North Africa. That is, we focus on regions where there is excess female mortality to begin with: in other parts of the developing world, such as East Asia (excluding China), Central Asia, and Latin America and the Caribbean we find no excess female mortality at adult ages.

To compute $\text{EFM}^1(a)$ and $\text{EFM}_e^1(a)$, we require data to determine $d^w(a)$, $d^m(a)$, $\sigma^w(a)$, $e^w(a)$, $\sigma^m(a)$, $e^m(a)$ and $\pi^w(a)$ for our regions of interest and for our reference region. Estimates of marital status by age group ($\sigma^w(a)$ and $\sigma^m(a)$) are available for all countries from the U.N. World Marriage Data (2012).[8]. The major sources of data on marital status used for these estimates are censuses, sample surveys and national estimates based on population register data or estimation methods using census data.

For estimates on population and mortality rates by age and gender ($d^w(a)$ and $d^m(a)$) we rely on the latest global, regional and country-level mortality rates by age and sex, released by the World Health Organization in 2016.[9] These estimates come from the Global Burden of Disease Study (GBD), which is the most comprehensive worldwide observational epidemiological study to date.[10] The GBD data is collected and analysed by a consortium of more than 2300 researchers in more than 130 countries. The data capture mortality rates in 195 countries by age and sex from 1990 to the present. The GBD engages a network of individual collaborators with expertise on allcause mortality; specific diseases, injuries, risk factors, and impairments; and countryspecific epidemiology. New data and methodological innovation lead to regular revisions of the estimates. The Core Analytic Team of the GBD is responsible for systematically collating and cataloguing data from sources such as vital registration, hospital data, disease registry data, surveillance systems, censuses, and large-scale household surveys. All available relevant sources of data for a given disease, injury, and risk factor and for allcause mortality are employed. For a country like India, there are a total of 1598 different sources used to generate the mortality estimates, whereas for countries in Sub-Saharan Africa, there are fewer sources, mainly due to less reliable vital statistic information, but still a very extensive number of sources are used. For example, the Burkina Faso estimates use 258 different sources. These include micro-level surveys conducted at the national level such as the: Demographic Health Survey; Continuous Multisectoral Survey; Core Welfare Indicators Questionnaire Survey; Demographic Survey; General Population Census; Health Statistical Yearbook; Household Health Coverage Survey; Malaria Indicator Survey; Multiple Indicator Cluster Survey; National Nutrition Survey; Performance, Monitoring, and Accountability Survey; Population and Housing Survey; Priority Survey; School Health Survey. Like India, the estimates

[8]Refer to: http://www.un.org/esa/population/publications/WMD2012/MainFrame.html

[9]Refer to: http://www.who.int/healthinfo/global_burden_disease/estimates/en/index1.html.

[10]http://thelancet.com/gbd.

also rely on information from a number of Health and Demographic Surveillance sites (HDSSs) as well as from more than 200 medical publications, which use first-hand data from locations in Burkina Faso. For Sub-Saharan Africa as a whole, the GBD estimates rely on information from 4889 different sources. [11] For all data sources identified, the sampling method, case definitions, and potential for bias are assessed. The GBD includes a Scientific Council, a Management Team, a Core Analytic Team, and a robust network of GBD Collaborators working together to produce the most accurate, uptodate, and comparable estimates of death burden worldwide. The entire GBD Study is under the leadership of the Principal Investigator Dr. Christopher Murray.[12] It is the case that the mortality estimates put together by the expert GBD teams entail some estimating assumptions based on certain epidemiological models. But these are country, regional, and disease specific. It seems highly unlikely that the mortality estimate differences between men and women that we observe in the data are universally biased, due to these modelling assumptions, so that we are systematically overestimating the extent of excess female mortality across developing countries.

Mortality rates by age, gender, *and* marital status (needed to compute $e^w(a)$ and $e^m(a)$) are more difficult to obtain as this information is typically not collected in any regular way by national statistical agencies. The U.N. Demographic Yearbook 2003 reports this information for all of these variables for several countries for which national vital statistics information was available. That gives us enough data to compute $e^w(a)$ and $e^m(a)$ for all developed countries for the year 2003. This information, however, is far rarer for developing countries. For our regions of interest, we have this data for four countries in Africa (Egypt, Mauritius, Reunion, and Tunisia), and nine countries in Asia (Hong Kong, Macao, Japan, Kazakhstan, Republic of Korea, Qatar, State of Palestine, Georgia, and Azerbaijan). We use this information from the two countries in North Africa (Egypt and Tunisia) to compute elevation ratios for the region of North Africa in 2003. We use the three countries from East Asia (Hong Kong, Macao, and Republic of Korea) to compute elevation ratio estimates for China and Southeast Asia in 2003. Likewise, we use the four countries in West Asia (Azerbaijan, Georgia, Palestine, and Qatar) to compute elevation ratios for that region. For our other regions of interest, there is no national vital statistics information available for death rates by marital status, with the exception of South Africa. The Department of Home Affairs in South Africa releases the number of registered deaths by marital status, age, and gender each year. We use this information to compute elevation ratios for the Southern Africa region in our analysis. For the remaining regions we turned to more micro-level studies. The INDEPTH Network, in particular, is an umbrella organization for a number of health and demographic surveillance sites (HDSSs) located in Africa and Asia.[13] Each HDSS collects detailed mortality data on a regular basis from specific, geographically defined, communities. This is an important source of information for countries that do not have functional civil registration and vital statistics systems. Deaths are followed up using verbal autopsy procedures (structured interviews with witnesses of the death, processed into cause-of-death information). INDEPTH has published a dataset covering over 100000 individual deaths across Africa and Asia. Each HDSS has been collecting health information on these populations for over two decades and information on mortality rates by marital status, age, and gender is regularly collected by all HDSSs but is not publicly available through this network - so we contacted each of the 47 HDDSs directly. Most of them were not willing to share this information, however, seven of them were. From this process, we are able to compute elevation ratios for East Africa (using information from two HDSSs, one in Uganda and the other in Malawi) and West Africa (using information from two HDSSs located in Burkina Faso and three HDSSs located in Senegal). To compute elevation ratios for Middle Africa, we aggregate the information we have from all of the HDSSs in both East and West Africa. None of the HDSSs located in South Asia were willing to share the required information. To compute elevation ratios in this region, we rely on the India Human Development Survey (collected by researchers from the University of Maryland and the National Council of Applied Economic Research (NCAER) in New Delhi). This is a nationally representative survey of 40,000

[11]A complete list of the input data sources into the GBD estimates can be found at http://ghdx.healthdata.org/gbd-2016/data-input-sources.

[12]Refer to http://www.healthdata.org/gbd.

[13]Refer to: http://www.indepth-network.org/.

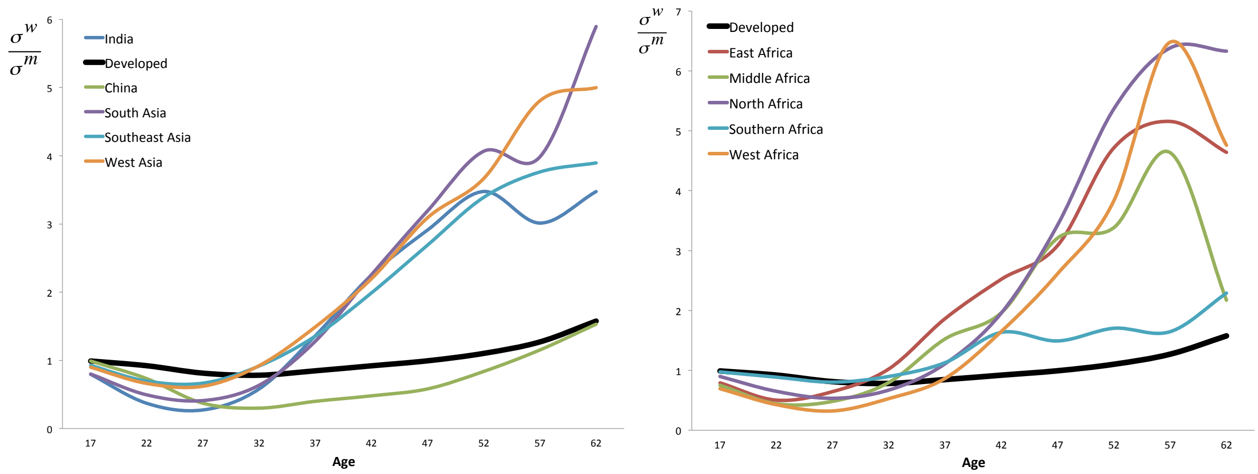
household across India, who were first interviewed in 2004 and then again in 2011. The survey collects information on household members who died between to the two rounds. Using this information we can compute estimated mortality rates by age, gender, and marital status. We use these estimates to compute elevation ratios for India and South Asia. We list all of our data sources in the Appendix.

In what follows we will compute our estimates of $\text{EFM}^1(a)$ and $\text{EFM}^1_e(a)$ for our regions of interest using the elevation factors we compute for these regions as just described. We do not have alternative region-specific estimates of elevation ratios at our disposal. Particularly, relying on the estimates from the HDSSs is somewhat problematic as they come from very small samples of data located in particular communities in small regions within a country and establishing external validity beyond its defined populations is difficult. This being said, a recent article by Byass (2016) compares the mortality estimates from the INDEPTH HDSSs and the GBD estimates and finds them to be highly congruent. Byass sates "Although the approaches, methods, and detailed inputs used were completely different, and independent, the high levels of concordance observed between them lend validity to both". INDEPTH sites are not purposefully designed to represent the countries in which they are located, and the assumption that findings from a localised site can be compared with estimates of national situations may be unjustified. However, what Byass finds, as might be expected, that countries with multiple sites had higher concordance correlations. In this sense having several distributed sites goes some way towards a national sample registration system. To this end, the fact that we have information from multiple HDSSs within specific countries in Africa (in particular Senegal and Burkina Faso) helps the validity of our estimates. Another weakness is using estimates from a nationally representative survey in India to capture elevation ratios in the rest of South Asia. One HDSS, located in Bangladesh, Matlab, would not share their data with us but they have published a set of earlier estimates of mortality rates my marital status (refer to Rahman 1993). Using these elevation ratios from the Matlab HDSS would increase our estimates of missing unmarried women substantially. In this sense, if anything our estimates are likely underestimates of the true number of missing unmarried women in the rest of South Asia.

For all of our variables we chose the data from the year 2005. The Global Burden of Disease (GBD) mortality estimates are available in five year intervals and hence this is the year that most closely matches our available data used to compute the regional elevation ratios. Given the paucity of data, it makes little sense to disaggregate "non-marriage" any further; on this, see Section 3.3. We focus on four different age categories, between 15 and 69, as defined in the GBD data.

Key to the computation of $\text{EFM}^1(a)$ is the size of $\theta(a)$, which is the female-male ratio of correction factors at age a in our region of interest, *relative* to the same ratio in the reference region. The larger the female-male correction factor in our region — influenced positively by both the elevation ratio and the overall incidence of non-marriage — the larger is the value of $\theta(a)$. That is, recalling that a typical correction factor $c(a)$ (dropping superscripts) is given by $\sigma(a)e(a) + (1 - \sigma(a))$, it follows that $\theta(a)$ is increasing in $e^w(a)/e^m(a)$ relative to the same ratio in our reference region, $\widehat{e}^w(a)/\widehat{e}^m(a)$. It is also increasing in $\sigma^w(a)/\sigma^m(a)$ relative to its reference value $\widehat{\sigma}^w(a)/\widehat{\sigma}^m(a)$. We now build these ratios from the data.

3.1. **Marital Status by Age.** We first consider $\sigma^m(a)$ and $\sigma^w(a)$, the incidence of unmarried men and women respectively in age group a, in both our regions of interest and in our reference region. What is most relevant for the determination of $\text{EFM}^1(a)$ is the size of the ratio $\sigma^w(a)/\sigma^m(a)$ relative to the same ratio in our reference region, $\widehat{\sigma}^w(a)/\widehat{\sigma}^m(a)$. Figure 1 plots this relative ratio for our regions of interest, and for developed countries, our reference region. We see that the ratio $\sigma^w(a)/\sigma^m(a)$ for less developed regions is higher than for developed regions for ages older than 35. The only exception is China where this ratio is lower than for developed countries.



(a) Asia **(b)** Africa

Figure 1. The Ratio $\sigma^w(a)/\sigma^m(a)$ for Different Regions and Ages. Source: U.N. *World Marriage Data* (2012).

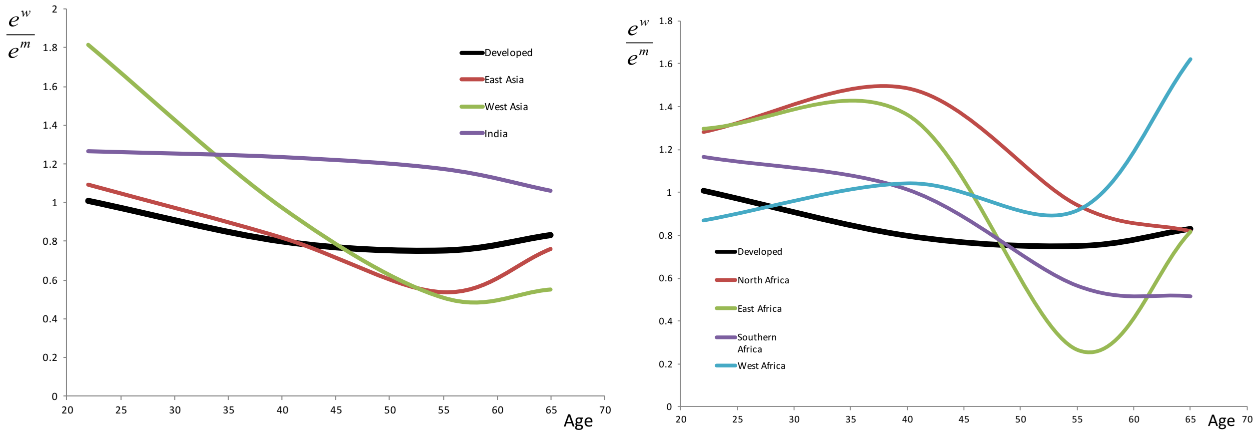
Unmarried individuals could be widowed, divorced or single. The second and third subcategories are largely symmetric across men and women, so the high values of $\sigma^w(a)/\sigma^m(a)$ are primarily driven by the proportions of widows (relative to widowers) in all regions. The fact that $\sigma^w(a)/\sigma^m(a)$ is lower in developed countries relative to developing is a sign of the fact that this imbalance is heightened in developing countries, at all age groups.[14] For more discussion, see Section 3.3.

3.2. **Elevation.** The second key component in our computation of EFM$^1(a)$ are the elevation factors $e^m(a)$ and $e^w(a)$, which reflect the relative mortality rates of unmarried and married individuals by gender and age.

As discussed, we construct eight main sets of elevation factors. The first is for our reference region, which is a population-weighted average across all developed countries. We then construct average elevation factors for North Africa (using the available data from Egypt and Tunisia); Southern Africa (using available data from South Africa); West Africa (using information from five HDSSs); East Africa (using information from two HDSSs); India (using information from the India Human Development Survey); East Asia (using data from Hong Kong, Macao, and Republic of Korea); and West Asia (using data from Azerbaijan, Georgia, Palestine, and Qatar).

What is particularly relevant for our computation of EFM$^1(a)$ is the size of the *ratio* of elevation factors — call it the *elevation ratio* $e^w(a)/e^m(a)$ — relative to the same ratio in our reference region, $\widehat{e}^w(a)/\widehat{e}^m(a)$. Figure 2 plots this elevation ratio for our eight different regions. We see that for all regions, in both Asia and Africa, this ratio is higher than in developed countries at all ages below 45-50 years old (there is one excepting for the youngest age category in West Africa). Moreover, this ratio is larger than 1 for most developing regions (in India at all ages and in Africa before the age of at least 40), implying that relative mortality rates for unmarried compared to married women are significantly higher than for men at these younger ages in Africa and South Asia.

[14] In addition, relative to developed countries, divorce is far less common in developing countries. In India and the rest of South Asia, the incidence of divorce is less than 1% in all age groups for men and women. In the rest of Asia it is at most 2%. In Africa, divorce is somewhat more common at around 5%.



(**a**) Asia (**b**) Africa

Figure 2. Relative Elevation Ratios by Age. Sources: U.N. Demographic Yearbook (2003), India Human Development Survey (2005, 2011), Iganga/Mayuge HDSS (Uganda), Karonga HDSS (Malawi), Kaya HDSS (Burkina Faso), Mlomp HDSS (Senegal), Nanoro HDSS (Burkina Faso), Bandafassi HDSS (Senegal), Niakhar HDSS (Senegal).

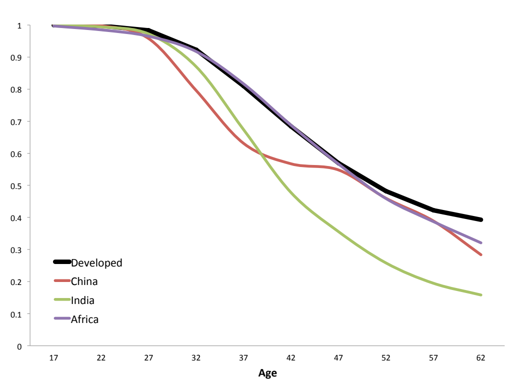
3.3. **Subcategories of Non-Marriage.** Both our methodology and computations treat the different unmarried categories (single, divorced, widowed) as one group. However, it is certainly the case that the proportions of individuals occupying each of these categories, as well as the corresponding elevation ratios across these different categories, varies by country.

Marriage occurs earlier in developing countries. It is therefore likely the proportion of the unmarried population that is single (as opposed to widowed or divorced), especially at younger ages, will be smaller in our regions of interest. In similar vein, because mortality rates are higher in developing countries, the proportion of widows (as opposed to those who are single or divorced) will be correspondingly higher in the regions of interest.

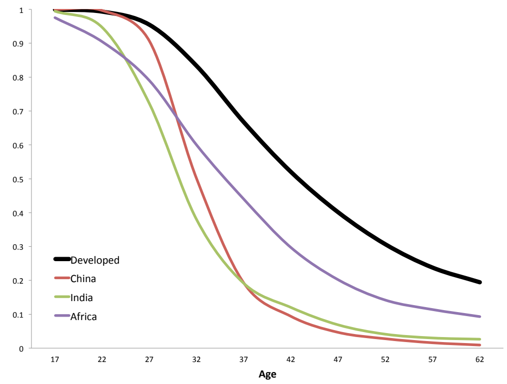
Figure 3 plots the proportion of unmarried individuals who are single or widowed in the different regions of interest, as well as in the reference region. Panels (a) and (b) show that the proportion of singles is higher in developed countries for all age groups, for both males and females. Panels (c) and (d) similarly show that the proportion of the unmarried who are widowed is lower in developed countries for all ages, again for both males and females. We see also that the proportion of widows is highest in India (for both men and women) compared to other developing regions.

Here is the basic reason that we cannot accommodate these finer subcategories in our analysis. Figure 3 shows that widowhood, for both men and women, is very uncommon in developed countries at younger ages. We therefore do not have reliable data on elevation ratios for this subcategory of the unmarried. Likewise, because divorce is very rare in some parts of the developing world, we do not have reliable data on elevation ratios for that subcategory. Furthermore, as already noted, even data on death rates conditioned on marital status is not universally available. Therefore, all things considered, we need to group the unmarried categories together in our analysis.

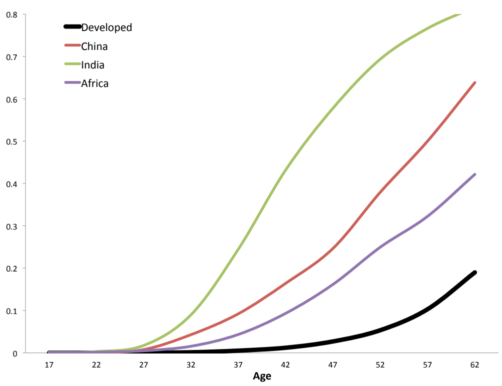
The question is whether our analysis overstates the case for unmarried missing women by lumping these subcategories together, as we are forced to do. Specifically, the concern is that at younger ages, we are more likely to be comparing widows in our region of interest to single (and divorced) people in our reference region, which could bias our estimates of $\text{EFM}^1(a)$ upwards. This would be the case if $\widehat{e}^w(a)/\widehat{e}^m(a)$ for



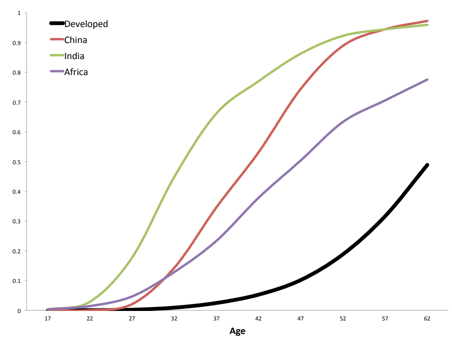
(a) Single/Unmarried: Males



(b) Single/Unmarried: Females



(c) Widowed/Unmarried: Males



(d) Widowed/Unmarried: Females

Figure 3. Proportions of Single and Widowed Individuals Among the Unmarried. Source: U.N. Demographic Yearbook (2003).

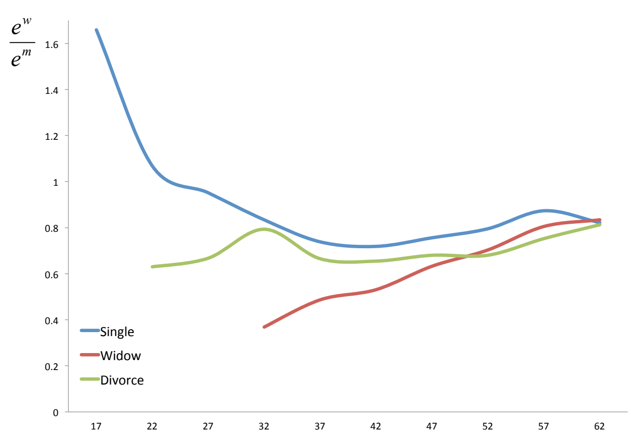


Figure 4. Elevation Ratios for Different Unmarried Subcategories. Source: U.N. Demographic Yearbook (2003).

widows were higher in our reference developed countries compared to the corresponding ratio for single individuals.

	Region									
	India		**S Asia**		**SE Asia**		**W Asia**		**China**	
Age	EFM^1	EFM^1_e	EFM^1	EFM^1_e	EFM^1	EFM^1_e	EFM^1	EFM^1_e	EFM^1	EFM^1_e
15-29	21	32	7	9	11	15	15	19	10	15
30-49	68	93	18	27	36	25	8	7	20	42
50-59	65	55	15	13	18	-7	0	-4	-6	-4
60-69	77	45	16	11	22	2	-11	-12	0	19
Total	**231**		**57**		**87**		**12**		**25**	

Table 1. Unmarried Excess Female Mortality (2005, in 000s): Asia. Sources: Global Burden of Disease Study (2015) U.N. World Marriage Data 2012; U.N. Demographic Yearbook 2003; India Human Development Survey (2005, 2011). EFM^1 refers to the number of missing women due to non-marriage and EFM^1_e refers to the number of missing women due to the marriage incidence effect.

Using the data available, we plot $\widehat{e}^w(a)/\widehat{e}^m(a)$ for the different unmarried categories for developed regions. We see, from Figure 4, that for all ages, the ratio $\widehat{e}^w(a)/\widehat{e}^m(a)$ is highest for single individuals, implying that if anything, the estimates to follow are biased downwards.

4. Unmarried Excess Female Mortality

With this information in hand, we can now compute $\text{EFM}^1(a)$ for the year 2005 in our regions of interest, for four different age categories between 15 and 69, using the methodology outlined in Section 2.

4.1. Asia.

We begin with Asia. As already noted, we concentrate on regions in which there is excess female mortality, so we do not compute $\text{EFM}^1(a)$ for East Asia (excluding China) and Central Asia.

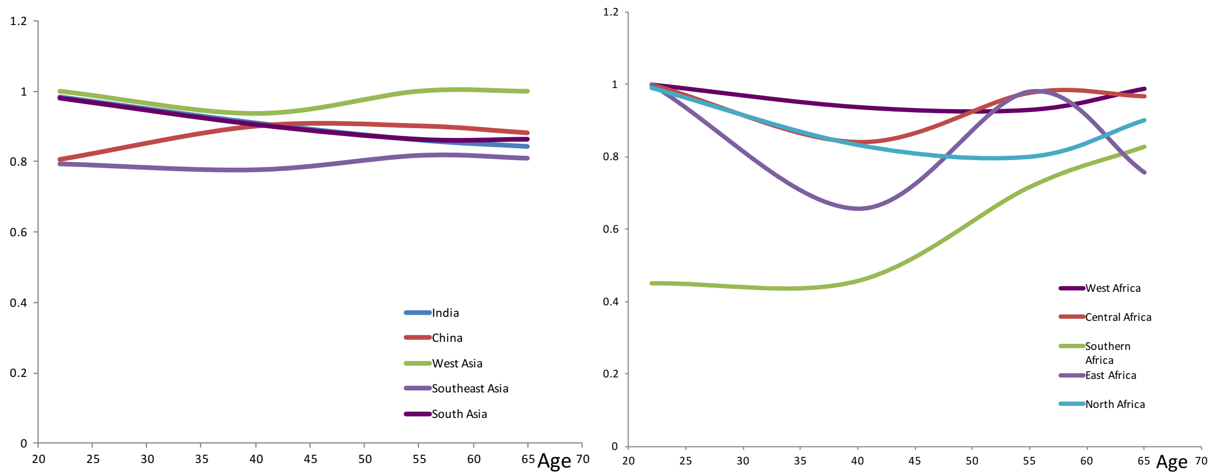
The first column of Table 1 lists the numbers of missing unmarried women in each age sub-category in the overall range 15–69, for India. In all we see that there are approximately 230,000 missing women due to non-marriage in India in the year 2005. Most of these women are over the age of 30 (90%).

The third column of Table 1 pertains to South Asia, excluding India. In this region, we see that there are close to 60,000 missing unmarried women each year. As a percentage of the total population of unmarried females, this is a lower number of missing women due to non-marriage in South Asia compared to India. A large proportion, 47%, are between the ages 30 and 49. Table 1 records 87,000 missing unmarried women each year in Southeast Asia. This number is larger than in South Asia, but as a percentage of the total population of unmarried females, there are fewer missing unmarried women in this region. Table 1 also provides our estimates for China and West Asia where we find relatively fewer missing unmarried women, compared to the other Asian regions. The few that are missing are at the younger adult ages (15 to 29).

In Section 2.3, we demonstrated how $\text{EFM}^1(a)$ could be divided into two components, the marriage incidence effect and the skewed elevation ratio effect. In Figure 5(a), we plot the proportion of $\text{EFM}^1(a)$ which can be attributed to the skewed elevation ratio effect for different Asian regions, by age. We see that $\text{EFM}^1(a)$ is primarily composed of this factor for all ages. This demonstrates the importance of a high value of

$$\theta(a) = \frac{c^w(a)/c^m(a)}{\widehat{c}^w(a)/\widehat{c}^m(a)}$$

in determining excess female mortality from the absence of marriage in Asia in general, and India in particular. That is, the female correction factor does indeed bear a higher ratio to its male counterpart in Asia,



(a) Asia **(b)** Africa

Figure 5. Proportion of EFM[1] attributable to the skewed elevation ratio effect. Sources: Global Burden of Disease Study (2015), U.N. World Marriage Data (2012), U.N. Demographic Yearbook (2003), India Human Development Survey (2005, 2011), Iganga/Mayuge HDSS (Uganda), Karonga HDSS (Malawi), Kaya HDSS (Burkina Faso), Mlomp HDSS (Senegal), Nanoro HDSS (Burkina Faso), Bandafassi HDSS (Senegal), Niakhar HDSS (Senegal).

relative to the reference region. As the discussion to follow will show, it is not just the elevation ratios but also the high rates of widowhood that drive the large values of $\theta(a)$.

Recall from Section 2.3 that the correction factor, which drives excess female mortality from the absence of marriage, is generated from two sources. One is that the elevation ratios $e^w(a)/e^m(a)$ are larger in our country of interest, compared to those in our reference region. The other is that the incidence of marriage is different across the two regions. We now attempt to disentangle these two effects. For each of the estimates in Table 1, we report corresponding estimates of $\text{EFM}_e^1(a)$ for each age group. Recall from Section 2.3 that this particular estimate of missing unmarried women allows for different elevation ratios but not different rates of marriage across the two regions. That is, we assume the rate of non-marriage by age group is identical across our region of interest and our reference region.

For India, we see (in the second column of Table 1) that these estimates are largest in the younger age categories (ages 15 to 49). This implies that the excess unmarried female mortality at these younger ages, primarily follows from the fact that the elevation ratio for India is high relative to the same ratio in our reference region. But above the age of 50, a different phenomenon takes over. Notice that $\text{EFM}^1(a) - \text{EFM}_e^1(a)$ is a measure of missing women due to changes in marriage incidence alone. These numbers are positive for ages above 50. So in other words, it is the large incidence of widowhood in India, at ages above 50, which is also driving the excess mortality for unmarried women at these older ages.

In South Asia, a similar pattern to India emerges. In the rest of Asia, it is also the case that the excess unmarried female mortality in the youngest age category (15 - 29), primarily follows from the fact that the elevation ratios are higher relative to the same ratio in our reference region. However, by comparison to South Asia, the larger incidence of widowhood, explains a larger component of the missing unmarried women in the rest of Asia.

4.2. **Africa.** We now turn to Africa.

	Region									
	East		**Middle**		**Southern**		**West**		**North**	
Age	EFM^1	EFM^1_e	EFM^1	EFM^1_e	EFM^1	EFM^1_e	EFM^1	EFM^1_e	EFM^1	EFM^1_e
15-29	15	18	0	-1	22	22	-8	-17	4	4
30-49	124	121	18	20	45	35	25	36	12	14
50-59	-6	-20	1	-1	2	0	10	6	11	5
60-69	20	7	2	1	-3	-4	2	5	4	0
Total	153		21		65		30		30	

Table 2. Unmarried Excess Female Mortality (2005, in 000s): Africa. Sources: Global Burden of Disease Study (2015) U.N. World Marriage Data 2012; U.N. Demographic Yearbook 2003; Iganga/Mayuge HDSS (Uganda); Karonga HDSS (Malawi); Kaya HDSS (Burkina Faso); Mlomp HDSS (Senegal); Nanoro HDSS (Burkina Faso); Bandafassi HDSS (Senegal), Niakhar HDSS (Senegal). EFM^1 refers to the number of missing women due to non-marriage and EFM^1_e refers to the number of missing women due to the marriage incidence effect.

Table 2 shows that there are approximately 300,000 missing unmarried women in Africa in the year 2005. Roughly 51% of these are from East Africa, where as a percentage of the total population of unmarried women the numbers are also the highest. By contrast, the countries of Middle and Lower Africa have the lowest levels of excess female mortality from non-marriage.

In Figure 5(b), we plot the proportion of $\text{EFM}^1(a)$ which is attributable to the skewed elevation ratio effect for the different African regions by age. We see that, like the regions of Asia, $\text{EFM}^1(a)$ is primarily composed of this component for all ages, except for Southern Africa. But overall, Figure 6 is in accordance with Figure 5 in that they both demonstrate the importance of a high value of $\theta(a)$ in determining excess female mortality from the absence of marriage in Africa as well. That is, the female correction factor does indeed bear a higher ratio to its male counterpart in Africa, relative to the reference region.

As seen from Table 2, the majority (75%) of the missing unmarried women across Africa are between the ages 30 and 49. From our computations of $\text{EFM}^1_e(a)$ across the regions of Africa in Table 2, we see that these are mainly due to the high elevation ratios in theses regions of Africa relative to the same ratio in our reference region.

5. Unmarried Excess Female Mortality and Missing Women

We are now in a position to compare our estimates of unmarried excess female mortality, $\text{EFM}^1(a)$, to estimates of overall excess female mortality, $\text{EFM}(a)$. In Table 3, we compute both kinds of excess mortality, $\text{EFM}(a)$ and $\text{EFM}^1(a)$, across the different age groups. Comparing these totals enables us to determine how much of the overall excess female mortality in developing regions can be attributed to not having a husband.

Across Asia and Africa, there are approximately 711,000 missing unmarried women in the year 2005. If we break this up by age, 52% of the missing women aged 30 to 49 are due to the absence of marriage, and 34% for the older ages (50 to 69). These estimates also vary by region, with India contributing the largest total number of missing unmarried women, followed by the countries of East Africa. China, West Asia, West Africa, Middle Africa, and North Africa all respectively make up less than 5% of the total missing unmarried women.

By comparing the first and second columns in each age group, we can determine what proportion of the total missing adult women are due to not having a husband. We see that this proportion in highest, at

	Age							
	15-29		**30-49**		**50-59**		**60-69**	
Region	EFM	EFM[1]	EFM	EFM[1]	EFM	EFM[1]	EFM	EFM[1]
India	234	22	183	68	139	65	261	77
South Asia	54	7	63	18	37	15	65	16
Southeast Asia	46	11	78	36	39	18	51	22
West Asia	15	15	21	8	12	0	21	-11
China	46	10	129	20	118	-6	202	0
East Africa	117	15	224	124	53	-6	46	20
Middle Africa	42	0	54	18	17	1	21	2
Southern Africa	38	22	70	45	11	2	5	-3
West Africa	111	-8	145	25	47	10	66	2
North Africa	16	4	22	12	15	11	22	4
Total	**720**	**97**	**991**	**374**	**490**	**110**	**760**	**130**

Table 3. Unmarried Excess Female Mortality (2005, in 000s). Sources: Global Burden of Disease Study (2015) U.N. World Marriage Data 2012; U.N. Demographic Yearbook 2003; India Human Development Survey (2005, 2011), Iganga/Mayuge HDSS (Uganda); Karonga HDSS (Malawi); Kaya HDSS (Burkina Faso); Mlomp HDSS (Senegal); Nanoro HDSS (Burkina Faso); Bandafassi HDSS (Senegal), Niakhar HDSS (Senegal). EFM refers to the number of missing women and EFM[1] refers to the number of missing women due to non-marriage.

approximately 40%, in the age category of 30 to 49. By comparison far less (13%) of the overall excess female mortality in the youngest age category (15 - 29) is explained by not being married.

Not being married explains the largest proportion of missing women in India, , approximately 40% of the missing adult women are due to not having a husband. By comparison very little (13%) of the excess female mortality in the youngest age category (15 - 29) is explained by not being married. In the older age categories (50 to 69) it is approximately 20%.[15]

6. Discussion

We are absolutely explicit that our analysis is fundamentally an accounting device that exposes different components of excess female mortality — non-marriage, in this particular instance — and contextualizes those components relative to overall excess mortality. That said, the analysis is incomplete without some discussion of the relevant pathways; specifically, the possible bias created by selection into marriage.

In the situation we study, there are three central differences between the region of interest and the reference region. The first is the usual scale effect: because our regions are poorer, they exhibit higher death rates for *both* men and women. The second difference is a corollary of these higher overall mortality rates: there is a

[15]In Table A.1 of the Appendix, we present results analogous to those in Table 3 but at the country level. We restrict ourselves to the sample of countries for which we have data on elevation ratios. Similar patterns emerge. The results for Bangladesh match the estimates for South Asia (in Table 3) as a whole, where 27% of the overall excess female mortality is explained by not being married. For countries within Africa we see that in Uganda and Malawi, 30% to 45% of excess female mortality is accounted for by non-marriage (the estimate is 35% for all of East Africa according to Table 3). The patterns for countries in Southern Africa and Northern Africa also match up. The estimate for South Africa in Table A.1 is 52%, whereas it is 53% for the region of Southern Africa in Table 3. Likewise, the estimates for Egypt and Tunisia in Table A.1 are on par with those for the region of North Africa in Table 3 (between 30% to 40%). Like the results in Table 3 for West Africa, we see in Table A.1 that very little of the excess female mortality is explained by non-marriage in Burkina Faso and Senegal, less than 10%.

higher incidence of "non-marriage" (via widowhood and widowerhood) in our regions of interest.[16] Finally, there is a gender bias, reflected in varying elevation factors across women and men, with the resulting relative elevation ratio then varying across the reference region and the region of interest.

6.1. **The Scale of Mortality and Unbiased Death Rates.** Following standard practice (for an extended discussion, see Anderson and Ray 2010), the entire exercise presumes that the relative, age-specific death rates in the reference region *are* the "unbiased" death rates, and it is precisely the departures from those "reference rates" that generate our estimates. How appropriate is this choice? How do we know that gender-based death rates are not somehow "naturally" different at different levels of development?

This assumption is well-nigh impossible to test with available data. The best we could do is *presume* — at least for a sizable set of countries once poor, or poor countries today — that there is no gender discrimination, so that the relative rates in those countries are the "natural" relative rates. In our earlier work (Anderson and Ray 2010) we did use Latin American and Caribbean countries as an alternative reference group. Our estimates of excess female mortality for the age group 15–69 decrease by only 10%. So one might presume that the reference regions are not far off the mark to begin with.

But there is a deeper conceptual reason: the use of any reference group that does not replicate what we see in developed countries today runs the risk of burying important gender differentials under the cover of an "alternative benchmark." For instance, we could be labeling as natural and non-discriminatory the tendency for females to die *relatively more* in developing countries. We see absolutely no reason why this should be the case. In our opinion, it is far more satisfactory to presume that the "natural" relative death rates are indeed constant with development, and then to view every departure from that benchmark as *prima facie* cause for suspicion (though not as conclusive evidence).

6.2. **Incidence of Widowhood and Elevation Ratios.** Our principal focus is on differences in the incidence of marriage as well as in the relative elevation ratios. As we explain in Section 2.3, our methodology allows us to disentangle these two effects.

Approximately 34% of the missing unmarried women are in the older age groups 50 to 69. Our computations demonstrate that excess female mortality amongst this older unmarried group is driven mainly by a high relative incidence of widowhood in the regions of interest. It may well be that this marriage incidence effect is not linked directly to gender discrimination, and is just the outcome of age/gender mortality correlations with development. Further research is needed to identify exactly the sources generating the significant excess female mortality from the absence of marriage amongst older women in parts of Asia and Africa.

For the 52% of missing unmarried women that are of reproductive age (between 30 and 49 years old), we have demonstrated that it is the skewed elevation ratio effect that drives excess female mortality at these younger ages. Younger unmarried women are missing not because their death rates are elevated relative to their married counterparts — that is true of the reference region as well — but because the elevation *factor* for women (compared to that for men) is relatively high in the parts of Asia and Africa we've studied. This is plausibly due to limited access to resources and health care for women in this very socially marginalized group.

That said, we want to be extremely cautious in interpreting our results as providing direct evidence of discrimination.[17] We do not want to assert that our numbers — large though they might be — fully represent

[16]It is true that the incidence of marriage is typically higher in our regions, but this is outweighed by the higher mortality rates, especially at later ages.

[17]For instance, the dominance of the elevation ratio effect for younger women may come from the fact that the ages of their children are relevant. Younger widows are likely to have younger offspring, who are less able to help look after them. To be sure, the fact that this skews the relative elevation rather than simply raising the correction factors by the same proportion for each gender continues to be suggestive of some form of female discrimination.

overt (or even implicit) discrimination against unmarried women. In general, there will be an entire complex of social, behavioural, and economic pathways that will need to be invoked. Our objective in this paper is simply to flag non-marriage as a factor in determining excess female mortality at older ages in a unified and comparable way across developing countries. The numbers are striking: in each year, over 710,000 unmarried women are missing.

6.3. **Selection Versus Protection.** The very last of the three central differences enumerated above, which is the variation of the elevation ratio across the reference region and the region of interest, will now be examined more closely. The magnitude of any *causal* link between marriage and excess unmarried female mortality is predictably hard to establish. In the literature examining the relationship between marital status and mortality in developed countries, a long-standing debate attempts to disentangle the role of "marriage protection" versus "marital selection" in explaining the observed differences (Goldman 1993, Sullivan and Fenelon 2014). Much of that literature focuses on the role of "marriage protection", i.e., the social, psychological, economic, and environmental benefits associated with having a spouse, that help to prevent premature mortality. When those protections are removed, death rates spike. Bereavement is an important reason, but it is not the only one. In developing countries, widowed women can suffer the vulnerabilities of reduced social support (Kochar 1999, Chen and Drèze 1995a,b).

A competing explanation is the role of "marital selection." For instance, if there is marriage sorting by economic or health status, the death rates of partners could be co-vary via socio-economic factors that influence mortality (Bowling 1987). It is also possible for selection to occur via re-marriage as healthier individuals are more likely to remarry (Sullivan and Fenelon 2014). This competing explanation is of particular interest in our context, as it runs against the spirit of our exercise, which is to suggest that widowhood could *precipitate* excess female mortality. If selection is the dominant consideration that drives a high elevation factor, then we could be attaching excessive importance to the widowhood argument.

6.3.1. *Relative Importance of Protection and Selection.* Obviously, it is practically impossible to conduct a randomly controlled experiment to identify the relative importance of protection and selection. Researchers have typically turned to large-scale individual-level longitudinal data to establish a causal role from marriage to better health outcomes, controlling for observables most likely linked to selection into marriage. For instance, Rendall et. al. (2011) conclude that these more recent analyses have led to a stronger case for the "marriage protection" hypothesis:

"Positive selection on unobserved characteristics cannot be ruled out as at least a partial explanation for our findings, and further efforts to model such selection are encouraged. However, the consistently negative coefficients for the unmarried of our study after controlling for observed socioeconomic and disability variables are themselves stronger evidence in support of a marriage protection effect . . . "

Moreover, as there does not appear to be significant differences in mortality rates for the different categories of "non-marriage" (never married, divorced/separated, and widowed) it is difficult to argue that the particular traits systematically determining worse health outcomes could simultaneously explain selection into these different categories. Indeed, there appears to be general agreement that selection factors do not come close to explaining all of the rise in spousal mortality; see, e.g., Allison and Christakis (2006), Boyle, Feng and Raab (2011), Espinosa and Evans (2008), Martikainen and Valkonen (1996), and Waldron, Hughes and Brooks (1996). In line with Rendall et al. (2011), Sullivan and Fenelon (2014) observe that "prior research on this subject indicates that selection is not the most important explanation for the association between widowhood and subsequent mortality." Their best estimate is that "approximately one third of the increase [in mortality rates] can be attributed to selection." Nevertheless, we need to examine this issue more closely.

6.3.2. *Accounting for Selection.* How might selection affect the findings of our exercise? Recall our basic equation that describes and decomposes excess female mortality at any age, reproduced here with the age argument removed for ease in writing:

$$\text{EFM}^1 = \left[1 - \frac{1}{c^w}\right]\text{EFM} + [\theta - 1]r^w \frac{\pi^w}{c^w}. \tag{10}$$

where θ is given by:

$$\theta = \frac{c^w/c^m}{\widehat{c}^w/\widehat{c}^m}. \tag{11}$$

6.3.3. *Selection and the Elevation Ratio θ.* There are two levels at which selection might bias our estimates. The first has to do with the "true" correction factor c^w once selection is introduced. Second, there is the effect on the elevation ratio θ. The latter effect is a bit more complex and we discuss that first before combining both factors. As far as the elevation ratio is concerned, then, the magnitude of the effect we estimate is threatened if women in developing countries experience a *larger* selection effect relative to their male counterparts than they do in the reference region. In that case, by (11), the term θ that we compute would be biased upwards relative to its true value, thereby increasing our estimates of excess female mortality stemming from non-marriage.

For instance, it must be the case that marital sorting on health status in developing countries is more severe relative to the corresponding sorting in developed countries, *and that this severity impinges more heavily on women*. Moreover, given our estimates, such an argument would further require that this "double-skew" of the sorting pattern is more prevalent in India (and East Africa) compared to other regions of Africa and China.

This is a subtle argument. The assertion that mortality heterogeneity in developing countries is an order of magnitude higher, or is more steeply related to income, may well be correct but it does *not* deliver the required implication. As an example, Ram et al. (2015) observe that in 2014 India, "the roughly 10-year survival gap between high-mortality and low-mortality districts was nearly as extreme as the survival gap between the entire Indian population and people living in high-income countries."[18] If there is spatial clustering in marriage, which is a reasonable position to take, it will lead to substantial selection effects, quite plausibly higher than those in developed countries. But *skewed* selection is not implied. For that, one would need to examine the *ratio* of female to male mortality, and how that affects the corresponding elevation ratios that are used to construct θ.

Table 4 provides the required computations based on data from Ram et al. (2015). They divide the districts of India into three mortality categories. The columns marked "m" and "w" show death rates per thousand (for men and women respectively), and these do rise sharply across the three district groupings. The *ratios* of these death rates by gender (shown in the columns marked "w/m") are less variable, however, and are suggestive of the possibility that there is little skew implied by selection *per se*. To confirm this, we compute the implied elevation factors for men and women, marked $\hat{e}^m$ and $\hat{e}^w$, that would be generated by these numbers. For instance, $\hat{e}^w$ would be calculated by first solving for the death rate of a widow were she to be randomly drawn from any of the districts. If all women marry men from the same district (which is the strongest case for selection in this example), this death rate would be high compared to the average death rate for women in all districts. And indeed it is high, but there is no skew across gender.[19] For comparison,

[18] We are grateful to an anonymous referee for bringing this reference to our attention, and for pushing us to discuss the implications of spatial heterogeneity in mortality.

[19] Women tend to marry younger than men in India so that the selection effect could happen at different ages for men and women. We've ignored that issue here. For it to change anything, risk selection into marriage be based not just on mortality levels but also systematically on mortality-age profiles; i.e., differential selection based on *how mortality changes with age*. If, for instance, all selection is based on mortality levels leaving mortality ratios across different ages unaffected, *relative* elevations will remain unchanged.

	District Mortality Type									Relative Elevations	
	Low			Medium			High			Implied	Actual
Age	m	w	w/m	m	w	w/m	m	w	w/m	$\hat{e}^w/\hat{e}^m$	e^w/e^m
15–29	1.3	1.1	*0.9*	1.8	1.7	*0.9*	2.5	2.3	*0.9*	**0.993**	**1.266**
30–69	7.8	5.0	*0.6*	11.0	7.5	*0.7*	15.7	11.1	*0.7*	**0.991**	**1.870**

Table 4. Relative Gender Mortality Across Low- and High-Mortality Districts of India, 2014. *Sources and Notes.* Columns m and w depict mortality rates per 1000 from Table 1 of Ram et al. (2015). Columns w/m plot the relative mortality rate. We compute implied elevation factors if it is assumed that *all* elevations come from selection into within-district marriage. For instance, $\hat{e}^w$ calculates the relative elevation ratio that would occur were a widow to be randomly drawn from any of the districts, compared to the average death rate for women in all districts. The populations used to weight the death rates are also taken from Ram et al. (2015); they are 137m, 148m and 143m (men 15-69) and 121m, 175m and 112m (women 15-69) for low-, medium- and high-mortality districts respectively. Column $\hat{e}^w/\hat{e}^m$ reports these relative implied elevation ratios and the final column reports the actual relative elevation ratios, e^w/e^m, used in our analysis.

we include the relative elevation factors that we used for our analysis in the last column, which indeed are significantly skewed across gender.

This discussion also suggests that the skew induced by selection is smaller than the observed ratio of correction factors, so that *the selection-adjusted ratio of correction factors is likely to be even larger than the observed ratio*. We use this information in the next section.

6.3.4. *An Overall Assessment of the Importance of Selection.* Of course, equation (10) also reminds us that quite apart from the issues with changing θ, a higher death rate from non-marriage also contributes to excess female mortality. This is the first term in equation (10): the larger the correction factor, the larger is our estimate, and if some of that increase in the correction factor comes from selection, our estimates of excess female mortality from non-marriage could be biased upward.

Suppose, then, that for the region of interest, a component s is equally responsible for elevated death rates (under non-marriage) for both men and women. Recalling (5), we see that

$$c^j = (1-\sigma^j) + \sigma^j(e^j_* + s) \tag{12}$$

for each gender $j = w, m$, where e^j_* stands for the elevation free of selection. For the reference region, denote the corresponding selection component by λs, where $\lambda = 1$ if selection is equally strong in both regions. So we have

$$\widehat{c}^j = (1-\widehat{\sigma}^j) + \widehat{\sigma}^j(\widehat{e}^j_* + \lambda s) \tag{13}$$

again for the two genders w and m.

Now differentiate EFM^1 with respect to s in (10). Using (12), this yields

$$\frac{d\text{EFM}^1}{ds} = \frac{[\text{EFM} - \text{EFM}^1]\sigma^w}{c^w} + \frac{d\theta}{ds}\frac{r^w\pi^w}{c^w}. \tag{14}$$

To calciulate $d\theta/ds$ we differentiate (11). Using (12) and (13), we obtain, after some manipulation:

$$\frac{d\theta}{ds} = \frac{\frac{c^m\sigma^w - c^w\sigma^m}{(c^m)^2} - \lambda\theta\frac{\widehat{c}^m\widehat{\sigma}^w - \widehat{c}^w\widehat{\sigma}^m}{(\widehat{c}^m)^2}}{[\widehat{c}^w/\widehat{c}^m]}. \tag{15}$$

Combining (14) and (15), we see that

$$\frac{d\text{EFM}^1}{ds} = \frac{[\text{EFM} - \text{EFM}^1]\sigma^w}{c^w} + \frac{r^w\pi^w}{c^w}\frac{\frac{c^m\sigma^w - c^w\sigma^m}{(c^m)^2} - \lambda\theta\frac{\widehat{c}^m\widehat{\sigma}^w - \widehat{c}^w\widehat{\sigma}^m}{(\widehat{c}^m)^2}}{[\widehat{c}^w/\widehat{c}^m]}. \tag{16}$$

To interpret this derivative, evaluate it for small s. In that case EFM^1 is at its true selection-free value. The derivative then tells us how the estimated value we report moves away from the true value as we introduce some selection. If the derivative is positive, our estimates are biased upward; if negative, then downward.

Everything in (16) is observed except for λ. However, the elevations in the reference region are of similar magnitude for men and for women. If anything, they are larger for men, as Figure 2 reveals. The same is roughly true of marriage rates in the reference region, though at advanced ages women are more likely to be widowed than men; see Figure 1. So as a first approximation, we can think of $\widehat{c}^w \simeq \widehat{c}^m$ and $\widehat{\sigma}^m \simeq \widehat{\sigma}^w$, which means that the value of λ is unimportant, and so (16) yields:

$$\begin{aligned}
\frac{c^w}{\pi^w}\frac{d\text{EFM}^1}{ds} &\simeq [\text{EFM} - \text{EFM}^1]\frac{\sigma^w}{\pi^w} + r^w\frac{c^m\sigma^w - c^w\sigma^m}{(c^m)^2} \\
&= \text{EFM}^0\frac{\sigma^w}{\pi^w} + r^w\frac{c^m\sigma^w - c^w\sigma^m}{(c^m)^2} \\
&= [d^w - \theta r^w]\,\frac{\sigma^w}{c^w} + r^w\frac{c^m\sigma^w - c^w\sigma^m}{(c^m)^2} \\
&\simeq \frac{d^w\sigma^w}{c^w} - \frac{c^w\sigma^m r^w}{(c^m)^2} \\
&= \frac{1}{c^w}\left[d^w\sigma^w - \theta^2 r^w\sigma^m\right], \qquad (17)
\end{aligned}$$

where the third line uses the formula for EFM^0 from (4), and the fourth and fifth line use $\theta \simeq c^w/c^m$, which follows from our approximation. Equation (17) gives us a condition for our calculations to be biased upward, which is

$$\theta^2 \simeq \left(\frac{c^w}{c^m}\right)^2 < \frac{d^w}{r^w}\frac{\sigma^w}{\sigma^m}. \tag{18}$$

Condition (18) is likely to hold. The right hand side is typically greater than 1 for each of the fractions, since $d^w > r^w$ — that's where the missing women come from — and it is also true that $\sigma^w > \sigma^m$, because rates of non-marriage for women are typically higher relative to men (due to higher rates of widowhood) in the region of interest. Indeed, they are significantly bigger; the ratio typically goes between 2 and 3. In contrast, on the other side we have c^w/c^m, which typically is not more than 1.1 or 1.2, and squaring it does not increase it by much.

Given that the bias is positive; that is, (18) is positive, then we must try to estimate the size of this bias. To address this, we have to make guesses about s. We can do this by using equation (12) and the existing literature, which places s at about 1/3 of the total elevation (Sullivan and Fenelon 2014). That is, if total elevation $e^w + s = 1.2$, then s should be around 1/3 of 0.2, which is 0.07. There may be a corresponding number for males but it seems reasonable to assume that the selection s should be the same — after all, the couple lives together and that should add the same number to their death rate. So, using (17), we see that the total correction is approximately:

$$\frac{d\text{EFM}^1}{ds}s = \frac{0.07\pi^w}{(c^w)^2}\left[d^w\sigma^w - \theta^2 r^w\sigma^m\right], \tag{19}$$

Region	**Age Range**				**Total**	**% of EFM**[1]
	15-29	30-49	50-59	60-69		
India	3.22	0.67	3.20	12.46	19.55	0.08
South Asia	0.52	0.26	0.83	2.81	4.42	0.08
Southeast Asia	0.73	0.53	1.58	4.04	6.88	0.08
West Asia	-2.39	0.19	0.62	2.27	0.70	0.06
China	0.78	-0.02	1.25	4.95	6.95	0.28
East Africa	1.22	1.16	2.73	2.18	7.29	0.05
Middle Africa	0.68	0.42	0.79	1.92	3.82	0.18
Southern Africa	0.39	0.30	0.34	0.61	1.63	0.02
West Africa	2.15	0.67	1.84	4.32	8.98	0.30
North Africa	-0.01	0.04	0.47	1.74	2.24	0.07

Table 5. Excess Unmarried Female Mortality from Selection Bias (2005, in '000). *Sources and Notes.* This table uses equation (19) to estimate the number of unmarried missing women (in '000) that can be attributed to selection. The last column expresses this number as a fraction of overall excess female mortality from non-marriage. Sources: Global Burden of Disease Study (2015), U.N. World Marriage Data 2012, U.N. Demographic Yearbook 2003, India Human Development Survey (2005, 2011), Iganga/Mayuge HDSS (Uganda), Karonga HDSS (Malawi), Kaya HDSS (Burkina Faso), Mlomp HDSS (Senegal), Nanoro HDSS (Burkina Faso), Bandafassi HDSS (Senegal), Niakhar HDSS (Senegal).

In Table 5, we compute the number of missing unmarried women in each age group that are due to selection bias, using (19). We see that, for example, in India, this total is 19,550 in the year 2005. This makes up approximately 8% of the total missing women due to non-marriage in India, which is the number reported in the final column of Table 5. Overall, summing across all age groups for all regions, the number of missing unmarried women due to selection bias is equal to approximately 62,000 out of our total, 712,000. That implies that our estimates are biased upwards by approximately 9% due to selection.

7. Conclusions

It is well known that the absence of marriage can pose significant risks, and that such risk can and does manifest itself in higher mortality rate for the unmarried. In principle, this is true of both men and women. There is a more subtle perception that the elevation of risk is higher for women then for men, and that this *relative* elevation is particularly acute for developing countries. This is the starting point of our paper, which attempts to establish the magnitude of this problem and situate it in a larger context: the phenomenon of "missing women" or excess female mortality in developing regions.

The numbers we put on this phenomenon are quite remarkable. All told, there are approximately 1 million missing women just between the ages of 30 and 49, each year. We find that 40% of these missing women of adult age — 400,000 of them — can be attributed to "non-marriage," which underlines the fundamental relevance and importance of this issue. Further research is needed to identify exactly the sources generating the significant excess female mortality for this very marginalized group of women.

References

Anderson, Siwan and Debraj Ray (2010) "Missing Women: Age and Disease," *Review of Economic Studies* **77**(4), 1262–1300.

Anderson, Siwan and Debraj Ray (2012) "The Age Distribution of Missing Women in India," *Economic and Political Weekly* **47**, No. 47-48, 87–95.

Byass, Peter (2016) ”Cause-specific mortality findings from the Global Burden of Disease project and the INDEPTH Network,” *Lancet* Nov;4(11):e785-e786.

Chen, Marty and Jean Dreze (1992) ”Widows and Health in Rural North India,” *Economic and Political Weekly* **27**, No. 43/44, WS81–WS92.

Chen, M. and J. Drèze (1995a), “Widows and Weil-Being in Rural North India,” in M. Das Gupta, T. N. Krishnan and L. Chen (eds.), *Women’s Health in India*, New Delhi: Oxford University Press.

Chen, M. and J. Drèze (1995b), “Recent Research on Widows in India: Workshop and Conference Report,” *Economic and Political Weekly* **30**, 2435–2450.

Farr, William (1858), “The Influence of Marriage on the Mortality of the French People,” in *Transactions of the National Association for the Promotion of Social Science*, edited by G.W. Hastings. London: John W. Parker. p. 504–13.

Goldman, Noreen (1993), “Marriage Selection and Mortality Patterns: Inferences and Fallacies,” *Demography* **30**, 189–208.

Jensen, Robert T. (2005) ”Caste, Culture, and the Status and Well-Being of Widows in India,” in *Analyses in the Economics of Aging*, David A. Wise (editor), University of Chicago Press, 357–373.

Johnson, N. J., E. Backlund, P.D. Sorlie, and C.A. Loveless (2000) ”Marital Status and Mortality: The National Longitudinal Mortality Study,” *Annals of Epidemiology* **10**(4), 224–238.

Kochar, A. (1999). “Evaluating Familial Support for the Elderly: The Intrahousehold Allocation of Medical Expenditures in Rural Pakistan,” *Economic Development and Cultural Change* **47**, 621–656.

Liu, Hui and David Johnson (2009), “Till Death Do Us Part: Marital Status and U.S. Mortality Trends, 1986–2000,” *Journal of Marriage and Family* **71**, 1158–1173.

Miguel, Edward (2005) ”Poverty and Witch Killing,” *Review of Economic Studies* **72**(4), 1153–1172.

Murphy, Michael, Grundy, Emily and Stamatis Kalogirou (2007), “The Increase in Marital Status Differences in Mortality up to the Oldest Age in Seven European Countries, 1990–99,” *Population Studies* **61**, 287–298.

Murray, C. J. L., Ferguson, B. D., Lopez, A. D., Guillot, M., Salomon, J. A. and O. Ahmad (2003), “Modified Logit Life Table System: Principles, Empirical Validation and Application,” *Population Studies* **57**, 165–182.

Nagata, C, N. Takatsuka, and H. Shimizu (2003) ”The Impact of Changes in Marital Status on the Mortality of Elderly Japanese,” *Annals of Epidemiology* **13**(4), 218–222.

Oppong, Christine (2006) ”Familial Roles and Social Transformations: Older Men and Women in Sub-Saharan Africa,” *Research on Aging* **28**(6), 654–668.

Rahman, Omar (1993) ”Excess Mortality for the Unmarried in Rural Bangladesh,” *International Journal of Epidemiology*, **22**(3), 445-456.

Rahman, Omar, Andrew Foster, and Jane Menken (1992) ”Older Widow Mortality in Rural Bangladesh,” *Social Science and Medicine*, **34**(1), 89-96.

Ram, Usha, Prabhat Jha, Patrick Gerland, Ryan J Hum, Peter Rodriguez, Wilson Suraweera, Kaushalendra Kumar, Rajesh Kumar, Rajesh Dikshit, Denis Xavier, Rajeev Gupta, Prakash C Gupta, Faujdar Ram, Age-specific and sex-specific adult mortality risk in India in 2014: analysis of 027 million nationally surveyed deaths and demographic estimates from 597 districts, The Lancet Global Health, Volume 3, Issue 12, 2015, Pages e767-e775, ISSN 2214-109X, http://dx.doi.org/10.1016/S2214-109X(15)00091-1 (http://www.sciencedirect.com/science/article/pii/S2214109X15000911).

Rendall, Michael S., Weden, Margaret M., Favreault, Melissa M., and Hilary Waldron (2011), "The Protective Effect of Marriage for Survival: A Review and Update," *Demography* **48**, 481–506.

Robards, Maria Evandroua, Falkinghama, Jane and Athina Vlachantonia (2012), "Marital Status, Health and Mortality," *Maturitas* **73**, 295–299.

Sen, Amartya (1990) "More than 100 Million Women are Missing," *The New York Review of Books*, **37**(20), 20 December.

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

Shor, Eran, Roelfs, David J., Bugyi, Paul, and Joseph E. Schwartz (2012), "Meta-Analysis of Marital Dissolution and Mortality: Reevaluating the Intersection of Gender and Age," *Social Science and Medicine* **75**, 46–59.

Sossou, Marie-Antoinette (2002) "Widowhood Practices in West Africa: The Silent Victims," *International Journal of Social Welfare* **11**, 201–209.

Subramanian, S.V., Felix Elwert, and Nicholas Christakis (2008) "Widowhood and Mortality among the Elderly: The Modifying role of Neighborhood Concentration of Widowed Individuals," *Social Science and Medicine* **66**, 873–884.

van de Walle, Dominique (2011) "Lasting Welfare Effects of Widowhood in Mali," World Bank Policy Research Working Paper 5734.

Waldron I1, Hughes ME, Brooks TL. (1996) "Marriage protection and marriage selection–prospective evidence for reciprocal effects of marital status and health." Soc Sci Med. Jul;43(1):113-23.

World Bank (2012), *Gender Equality and Development*, World Development Report.

Zheng, Hui and Patricia A. Thomas (2013), "Marital Status, Self-Rated Health, and Mortality: Overestimation of Health or Diminishing Protection of Marriage?" *Journal of Health and Social Behavior* **54**, 128–143.

APPENDIX

Data Sources.

Rates of non-marriage by Age and Sex: U.N. World Marriage Data (2012): http://www.un.org/esa/population/publications/WMD2012/MainFrame.html.

Mortality rates by Age and Sex: Global Burden of Disease Study (2005): http://www.who.int/healthinfo/global_burden_disease/estimates/en/index1.html.

Elevation Ratios for Developed Countries: U.N. Demographic Yearbook (2003): https://unstats.un.org/unsd/demographic/products/dyb/dybsets/2003%20DYB.pdf

Elevation Ratios for North Africa (Egypt and Tunisia): U.N. Demographic Yearbook (2003): https://unstats.un.org/unsd/demographic/products/dyb/dybsets/2003%20DYB.pdf

Elevation Ratios for East Asia (Hong Kong, Macao, and Republic of Korea): U.N. Demographic Yearbook (2003): https://unstats.un.org/unsd/demographic/products/dyb/dybsets/2003%20DYB.pdf

Elevation Ratios for West Asia (Azerbaijan, Georgia, Palestine, and Qatar): U.N. Demographic Yearbook (2003): https://unstats.un.org/unsd/demographic/products/dyb/dybsets/2003%20DYB.pdf

Elevation Ratios for Southern Africa: Department of Home Affairs in South Africa - Registered Deaths: https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/421

Elevation Ratios for India and South Asia: India Human Development Survey: http://ihds.info/

Elevation Ratios for East, West, and Middle Africa:

Karonga HDSS (Malawi): https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3396313/

Iganga/Mayuge HDSS (Uganda): http://igangamayuge-hdss.mak.ac.ug/index.php/en/

Mlomp HDSS (Senegal): http://www.indepth-network.org/member-centres/mlomp-hdss

Kaya HDSS (Burkina Faso): https://www.ncbi.nlm.nih.gov/pubmed/23918849

Nanoro HDSS (Burkina Faso): https://www.ncbi.nlm.nih.gov/pubmed/23045201

Bandafassi HDSS (Senegal): https://academic.oup.com/ije/article/43/3/739/2949552/Health-amp-Demographic-Surveillance-System-Profile

Niakhar HDSS (Senegal): https://www.ncbi.nlm.nih.gov/pubmed/24062286

Unmarried Excess Female Mortality by Country.

	Age							
	15-29		30-49		50-59		60-69	
Region	EFM	EFM[1]	EFM	EFM[1]	EFM	EFM[1]	EFM	EFM[1]
India	233903	21526	182876	68410	138991	64551	261162	76864
Bangladesh	18895	4437	26077	5002	13706	6143	27451	7733
China	46143	10409	129395	20195	118525	-5922	202142	175
Azerbaijan	254	185	304	211	125	98	907	-610
Uganda	9350	2678	19984	8552	4417	-1929	3813	1818
Malawi	7102	4207	19451	9052	3411	595	2517	1021
Burkino Faso	3678	848	6435	1346	1712	-163	2411	-511
Senegal	3160	-1839	2504	115	911	641	1691	625
South Africa	33565	19366	59334	37352	8789	1277	3443	-3287
Egypt	3181	1511	4549	3168	6189	2187	9278	-315
Tunisia	273	-32	745	275	248	124	291	138

Table A.1. Unmarried Excess Female Mortality by Country (2005). *Sources and Notes.* Global Burden of Disease Study (2015), U.N. World Marriage Data 2012, U.N. Demographic Yearbook 2003, India Human Development Survey (2005, 2011), Matlab HDSS (Bangladesh), Iganga/Mayuge HDSS (Uganda), Karonga HDSS (Malawi), Kaya HDSS (Burkina Faso), Mlomp HDSS (Senegal), Nanoro HDSS (Burkina Faso), Bandafassi HDSS (Senegal), Niakhar HDSS (Senegal). EFM refers to the number of missing women and EFM1 refers to the number of missing women due to non-marriage