Proximate Sources of Population Sex Imbalance in India

Emily Oster
University of Chicago and NBER
January 17, 2008

Abstract

There is a population sex imbalance in India. Despite a consensus that this imbalance is due to excess female mortality, the specific source of this excess mortality remains poorly understood. I use micro-data on child survival in India to analyze the proximate sources of the sex imbalance. I address two questions: when in life does the sex imbalance arise, and what health or nutritional investments are specifically responsible for its appearance. I present a new methodology, which uses microdata on child survival and explicitly takes into account both the possibility of naturally occurring sex differences in survival, and possible differences between investments in their importance for survival. I find significant excess female mortality in childhood, particularly between the ages of 1 and 5, and argue that the sex imbalance that exists by age 5 is large enough to explain virtually the entire imbalance in the population. Within this age group, sex differences in vaccinations explain between 20 and 30% of excess female mortality, malnutrition explains an additional 20% and differences in treatment for illness play a smaller role. Together, these investments account for approximately 50% of the sex imbalance in mortality.

1. Introduction

India has a serious population sex imbalance. There are around 108 men for every 100 women in the country as a whole. In a country with the same level of development and "normal" mortality patterns, we would expect to see around 100 men for every 100 women. Sen (1990,1992) coined the phrase "missing women" to describe this population imbalance, and attributed it to sex discrimination. Consistent with this view, other authors (Visaria, 1971; Kishor, 1993) have argued, based on census data and other sources, that the sex imbalance is almost certainly due to excess female mortality.

There is a very large literature on the underlying sources of parental sex preferences (see, for example, Rosenzweig and Schultz, 1982; Agnihotri, 2000; Agnihotri et al, 2002; Murthi et al, 1995; Qian, 2007). These papers focus on the relative contributions of factors such as female labor force participation and female education in determining overall sex ratios. There is a second literature, more closely

related to this work, focusing on the proximate sources of female mortality.[1] That is, conditional on preferences, what specific treatments (or lack thereof) are responsible for the differences in mortality (Basu, 1989; Griffiths et al, 2002; Borooah, 2004; Pande, 2003; Mishra et al, 2004).

Despite this second literature, we still lack a coherent overall picture of the proximate sources of excess female mortality. This paper focuses on two primary questions: at what ages does most of the excess female mortality occur, and what is the relative contribution of various forms of neglect to this excess mortality. In contrast to most of the existing literature, I am concerned not only with whether various health and nutrition inputs play a role, but *how large* a role.

The methodology used here, formally outlined in Section 2, differs from most of the previous literature in two ways. First, I use data from Africa on sex differences in mortality and child health investments as a comparison for India. Existing literature (for example, Das Gupta, 1987) has often focused solely on sex differences in mortality in India. However, since boys are more likely to die in a world with equal treatment, the lack of a comparison group is likely to *understate* the extent of excess female mortality. Second, when considering the proximate sources of excess female mortality in childhood, I consider not only the difference in treatment, but also the importance of that treatment for mortality (i.e. the difference in mortality probability with and without treatment). Multiplying these two factors will give us full information about what the importance of each element is in understanding the overall excess female mortality. Existing literature generally considers only the difference across sexes in each treatment, and not the importance of these treatments in mortality, which is crucial for evaluating the relative contribution of each input (Basu, 1989; Griffiths et al, 2002; Pande, 2003; Borooah, 2004; Mishra et al 2004).[2]

In Section 4 I use micro-data (discussed in Section 3) to identify exactly the age source of the excess female mortality in childhood and to explore the importance of childhood sex bias in the overall imbalance.[3] The results here suggest that there are

[1]Throughout the paper I will refer to "proximate sources" of excess female mortality. I define a "proximate source" as an investment that differs across sexes. For example, if vaccination levels are higher for boys than for girls, vaccination is likely to be one "proximate source" of excess mortality.

[2]This should not necessarily be taken as a criticism of this work. Generally, the authors do not intend to calibrate the importance of different explanations in the sex bias, but rather to demonstrate that one particular explanation might play a role.

[3]Elsewhere (Oster, 2005) I argue that a naturally occurring higher sex ratio at birth resulting from hepatitis B can explain a fraction of the sex bias. However, in the case of India, this fraction is small –

important variations within young children. In particular, in all areas of India there is relatively little excess female mortality between the ages of a few months and 2 years, but substantial excess mortality between 2 and 5 years. In this section I also present evidence on the contribution of the under-5 sex ratio bias to the overall bias. Using demographers' life tables (Coale, Demeny and Vaughn, 1983) I calculate the expected sex ratio overall in India *if* we assume the empirically observed sex ratio at 5 years of age, and normal mortality thereafter. This exercise suggests that virtually all the sex ratio imbalance in the country can be explained by excess under-5 mortality.

Section 5 discusses the proximate sources of this excess female mortality between the ages of 2 and 5. Consistent with previous literature, I focus on biases in nutrition and medicine. The evidence here suggests that, contrary to some of the previous literature, sex differences in vaccinations play a very large role in the sex imbalance, explaining about 20 to 30 percent. Malnutrition explains about 20%. Interestingly, differences in treatment for respiratory infections and diarrhea together explain only around 5% of the imbalance. Around 50% is left unexplained by these childhood investments.

The results here have potentially important policy implications, suggesting that increases in vaccinations for girls could have a large impact on the overall sex imbalance in India.

2. Methodology

This section discusses the methodology used for estimating both the overall sex imbalance in mortality by age and the contribution of various investments to this sex imbalance. To illustrate the basic concept, define D as the differences between sexes in some investment (for example, the difference in the chance of measles vaccination). Define μ as the importance of this investment in mortality (for example, the difference in mortality probability if vaccinated and unvaccinated) and Ψ as the overall excess female mortality. The share of the overall difference that is explained by this investment is, therefore, simply

$$\frac{D\mu}{\Psi} \tag{1}$$

That is, the overall contribution is simply the expected excess mortality resulting from

only around 20% – implying that most of the bias is due to excess mortality. In this paper, I abstract away from a biased sex ratio at birth by using information on deaths for children ever born.

differences in this particular investment ($D\mu$) divided by the total excess mortality. The challenge, then, is estimating D, μ and Ψ.

First, consider the estimation of overall differences in mortality (Ψ from equation 1 above). This is an input to understanding the importance of different investments, but is also, when calculated for each age group, the parameter that will tell us the importance of each age group in the overall excess female mortality. This variable is, intuitively, the difference between actual and expected probability of death for girls. In other words, Ψ measures how much more likely a girl is to die relative to what we would expect based on mortality of boys.[4] Perhaps the most obvious way to estimate this would be to simply calculate the difference between male and female mortality in India and assume that Ψ is equal to that difference (in other words, to assume that what we expect is girls to have the same mortality as boys). The problem with this, however, is that there may well be differences between sexes even in non-discriminatory environments. If this is true, simply comparing the two sexes within India may understate (or overstate) the excess female mortality.[5] To solve this problem, I employ a "difference in difference" technique, in which I use data on India and a comparison region to evaluate the difference in mortality in India relative to the "expected" difference.[6] The equation estimated is:

$$\Pr(die) = (\alpha) + (\beta_0)(girl) + (\beta_1)(India) + (\beta_2)(girl \times India) + \Phi\mathbf{X} \qquad (2)$$

This regression relates the probability of death to child characteristics. $\Phi\mathbf{X}$ is simply a vector of controls – for example, mother's education, family income and other variables – that may affect child mortality. α is a constant in the regression. If $\mathbf{X} = 0$ (i.e. the value of all controls is equal to zero), then α is equal to the probability of death for a boy in the comparison region. The β coefficients measure differences in the probability of death across sex and area: β_0 is the difference in probability of

[4] It is worth noting, at this point, that the concept of an expected death rate based on male mortality presumes some standard mortality schedules to which the actual mortality can be compared. In practice, it is not possible to perfectly identify the biological expected mortality paths for men and women. Empirically, I will rely on data from a less discriminatory environment to provide information on the "expected" relative mortality for boys and girls. This is discussed more extensively in the data section.

[5] It is very frequently observed that men are more likely to die at all ages in non-discriminatory environments, but the reasons are not obvious. Wells (2000) provides good links to the literature on the existence of this effect, and presents one potential explanation.

[6] The comparison region used is Sub-Saharan Africa. This is discussed in more detail in the data section.

death for girls versus boys in the comparison region, and β_1 is the difference in probability of death (on average) between India and the comparison region.

The coefficient of interest is β_2, the interaction between being a girl and living in India. The coefficient on this interaction is the sex imbalance in mortality. If India is similar to the comparison region, then we should find $\beta_2 = 0$. If girls are disadvantaged, we should find $\beta_2 > 0$. In the language of equation 1, $\beta_2 = \Psi$.[7]

Equation 2 illustrates the problem with comparing death rates for boys and girls within India and using that comparison to measure excess female mortality. If we estimate equation 2 using only data from India, then $India = 1$ for all observations. In that case we will *not* be able to separately identify β_2 and β_0, and we will observe that the coefficient on $girl$, the Ψ we are interested in, is equal to $\beta_2 + \beta_0$. That is, we will not be able to separate the effect of sex overall from the effect of sex in India, and the coefficient measuring excess female mortality will not be interpretable as such.

The second methodological issue is identification of $D\mu$. I focus on two primary analyses: individual regression (which will estimate the entire quantity $D\mu$) and direct calculation of D and μ separately. Consider first the individual regression. Imagine that I have an individual-level panel in which I observe, the level of health investment and mortality outcomes for children. I can then estimate the quantity $D\mu$ by comparing the coefficient on $girl \times India$ in two difference in difference regressions – the first without controls for the health investment, and the second with these controls. In particular, denoting the health investment as Z, I first estimate equation 2 above, and then equation 3 below.

$$\Pr(die) = \delta + \gamma_0 (girl) + \gamma_1 (India) + \gamma_2 (girl \times India) + \gamma_3 (Z) + \Phi X \qquad (3)$$

Given these regressions, $D\mu = \beta_2 - \gamma_2$.

Perhaps the easiest way to see the intuition behind this calculation is to think of Z as an omitted variable in equation 2. β_2 captures the effect of many investments, one of which is Z. By not controlling directly for Z, β_2 is "upward

[7]It is important to note that, while I will continue to refer to this as a difference-in-difference regression, that does not connote anything about identification. What is done here is simply a mechanical adjustment for baseline differences between the sexes.

biased". Controlling for Z will decrease β_2, with the amount depending on how important Z is in explaining the mortality imbalance. More concretely, we model the relationship between Z and the interaction by equation 4 below:

$$Z = \eta + \upsilon_1 (girl) + \upsilon_2 (India) + \upsilon_3 (girl \times India) \quad (4)$$

We then note, based on the omitted variable intuition, that $\beta_2 = \gamma_2 + (\gamma_3)(\upsilon_3)$. This means that $\beta_2 - \gamma_2 = (\gamma_3)(\upsilon_3)$. From this it is straightforward to see why this is an estimate of $D\mu$: γ_3 is just a measure of the effect of health investment on mortality (μ) and υ_3 is a measure of the sex bias in that investment (D). The product of these two will give us the share explained by that particular investment. As noted, this analysis will require an individual-level panel dataset (or enough information to construct one).

A significant concern with this approach is that the elements of Z may "over-control" and soak up some of the effect of parental preferences. If Z measures vaccination, but differences in vaccination are simply a proxy for preferences and are perfectly correlated with all other forms of discrimination, then the difference between β_2 and γ_2 will capture much more than just the effect of vaccination. However, this will only be an issue if vaccination overall is correlated with parental sex preferences *and* if mortality is correlated with sex preferences, which does not seem to be the case empirically.[8] Regressing mortality and vaccination on the parental reported ideal sex ratio (parents are asked about their ideal number of sons and daughters in the later survey waves) yields insignificant and small coefficients (results available from the author). Despite this, we may still worry about omitted parental preferences. One way to partially adjust for possible preference differences is to include some simple preference controls – in particular, the mother's reported ideal sex ratio. When I do this, the results do not change. Of course, this control may not fully capture preferences and it remains a concern. One advantage of the second methodology discussed below is that these concerns will be largely avoided.

[8]To see this, consider the discussion of the omitted variable bias intuition in the text above. To say that we "over-control" implies that the adjustment between the two regressions is too large – that is, $(\gamma_3)(\upsilon_3)$ is too big. This will be the case if $\hat{\gamma}_3$ – the relationship between vaccination and mortality – is over-estimated. Based on the standard omitted variable bias arguments, omitting parental preferences will be a problem if measles vaccination is correlated with preferences and mortality is correlated with preferences.

The second methodology used to calculate $D\mu$ is direct estimation of D and μ. In particular, I first use the NFHS to directly calculate the sex differences in treatment. This is done by estimating equation (4) above – the estimate of D is simply υ_3. I then obtain estimates of μ from the existing literature, based on studies where mortality outcomes are observed for children with varying levels of health investment. There are two advantages to this approach. First, because the estimates of the effect of treatment on mortality come from other surveys, there is less concern about bias arising from this particular sample. Second, this technique will allow me to get estimates for the effect of nutrition and medical treatment, as well as for vaccination. As a final robustness check, I replicate the individual-level analysis using data at the regional level. Although this is likely to be the least appealing methodology, it does allow me to control for all of the elements of mortality simultaneously.

Before moving on to the data and results, it is worth briefly discussing how the methodology used here differs from that of the existing literature. There are two basic differences. First, most existing literature (for example, Das Gupta, 1987) uses only a difference approach – comparing the death rates of boys and girls in India. This will generally underestimate true excess female mortality because boys are more likely to die in non-discriminatory environments. Second, the existing literature on proximate sources of excess mortality generally focuses on estimating the differences in treatment by sex (i.e. D from the discussion above) and not the effect of these treatments on mortality (Basu, 1989; Griffiths et al, 2002; Pande, 2003; Borooah, 2004; Mishra et al 2004). Without adjusting for differences in μ it is very difficult to say anything conclusive about which inputs are more important in explaining the sex differences.[9]

3. Data

The analyses here will be run using individual-level microdata on child survival and health investments.[10] For India, the data used are from two waves of the National

[9] In the existing literature there is also a lot of focus on the differences between North and South India. If I separate into the two regions, I find that sex imbalances are higher in North India in virtually all of the inputs and in excess mortality, but the conclusions about patterns by age in childhood and about which proximate sources are most important will hold.

[10] This is in contrast to much of the literature on this topic, which relies on district-level data on sex

Family and Health Survey (NFHS) (1992-1993 and 1998-1999), which covers approximately 90,000 women in each wave. Women are asked about their birth history, including children ever born, dates of birth, if the children are alive and, if not, when they died. In addition, for children under five, information is collected on vaccination, medical treatment and malnutrition. In the 1992-1993 survey, vaccination information was collected for all children, including those who had died. In the 1998-1999 survey, however, this information was not collected for children who had died. The analyses of vaccination, therefore, include only the 1992-1993 round of the NFHS.

As discussed in the methodology section, the size of the sex imbalance in mortality and investments in India is evaluated relative to the size of this imbalance in a comparison area. This will allow me to difference out any differences across sexes (favoring boys or girls) that occur in apparently non-discriminatory (or less discriminatory) environments. The literature on the "missing women" suggests two natural comparisons: Sub-Saharan African (Sen, 1990; Sen, 1992) and demographer's life tables (Coale, 1991; Klasen and Wink, 2002). Sex differences in mortality in Sub-Saharan Africa are similar to those predicated in the life tables, suggesting that either comparison will give similar results. The advantage of using Sub-Saharan Africa, as I do here, is that the same type of microdata on children is available from a number of countries. The Demographic and Health Surveys (DHS) in Africa mirror the NFHS, so the difference-in-difference analysis can be run at the level of the individual child. The comparison countries are Kenya, Namibia, Zambia, Tanzania and Zimbabwe.

The three child investments analyzed are vaccination, malnutrition and treatment for disease. There are seven possible vaccinations (three DPT vaccines, two polio vaccines, a measles vaccine and a BCG vaccine). In general, I will use two measures of vaccination: the total number of vaccinations reported by the mother and the total number marked on the child's health card. The results are extremely similar if I include dummies for each vaccination.

Information on malnutrition is based on actual height and weight measurements. Living children under four in each household are measured and weighed. Their percentile weight-for-age is reported (weight-for-height and height-for-age are also reported, and all are very closely linked). I define children as severely

---

ratios. The advantage of using the individual-level data is that we observe directly the relationship

malnourished if their percentile weight-for-age is less than 60% of the reference median for their age and sex. I use this indicator rather than a continuous measure since research on the effect of malnutrition on mortality indicates that mortality is largely unaffected by malnutrition above 60% of the reference median, but increases sharply below that (Chen et al, 1980).

To evaluate differences in medical treatment, parents were asked whether each of their (living) children under four had diarrhea or symptoms of a respiratory infection in the last two weeks. If the answer was yes, they were asked what treatment was provided. I report children as having been treated if their parents report having given the child any treatment (including doctor visit, home remedies, etc). In these data, differentiating by treatment type has little effect on the sex difference.

4. Age-Specific Origins of Excess Female Mortality

The first set of results estimate the baseline excess female mortality. I estimate equation (2) for age groups ranging from birth to ten years. The dependent variable is a series of indicators for having died within a particular age group. For example, the first variable is a 0-1 dummy for whether a child born in the last ten years died before the age of six months; the second variable is a 0-1 dummy for whether the child died between six months and one year, *conditional on* having lived to six months. The additional age groups are one-two years, two-four years, four-six years, six-eight years and eight-ten years.

The results of this analysis can be seen graphically in Figure 1, which plots actual and expected mortality for girls in India by age group, with the expected mortality based on male mortality in India and the sex difference in Africa. By the age of ten, the actual probability of female deaths is almost 12%, compared with an expected probability of slightly less than 10%. Nearly all of this imbalance seems to arise between the ages of one and four, when expected mortality is around 1.4% and actual mortality is a full 2.4%. The regression analog to this figure appears in Table 1, where the difference-in-difference estimate is the coefficient on $Girl \times India$. Consistent with the picture, the difference is statistically significant between six months and six years, but not in the youngest or oldest groups.

Controls in this regression include child age, maternal age, maternal

---

between health investments and mortality.

education, birth order dummies and total number of siblings. In general, these enter with the expected sign and are unremarkable. However, the control for total number of siblings is worth a brief discussion. It is frequently suggested that one of the major reasons why female mortality is higher is that girls are, on average, in larger families (due to some form of sex-biased stopping rule). This seems to be somewhat true, as excluding the control for number of siblings leads to a larger estimate for the interaction between sex and India. In this sense, I can say that one “proximate source" of excess mortality is larger family size. However, exactly what health investments are denied in larger families remains to be analyzed.

The results in Table 1 give a sense of the magnitude of excess female mortality in childhood and the periods of childhood which are most crucial. A related question is how important childhood is in explaining the overall sex imbalance. To get a sense of this issue I calculate the predicted sex ratio in the population (based on life tables) taking the sex ratio at age five as given. If the predicted sex ratio in the population based on this calculation is much lower than the actual sex ratio, then it suggests that any excess mortality up to age 5 is probably unimportant in the overall sex bias. If, in contrast, the predicted and actual sex ratios are similar it would suggest that excess female mortality before age five explains most of the overall sex imbalance.

The result of these calculations appears in Table 2 (details of the calculation are in Appendix A) and the results suggest that a very large share of the sex bias can be explained by events occurring up to age five. This, in turn, suggests that understanding the proximate sources of mortality in this age bracket may go far in helping us understand the overall problem. This is not surprising. Mortality rates among young children are much higher than among prime-age adults so we would expect mortality in childhood to contribute to a large share of the sex imbalance simply because the level is higher. It is worth noting, however, that the relationship is not mechanical. Even though mortality from zero to six months is much higher than mortality later in childhood, that period does not contribute very much to the sex imbalance.

5. Proximate Causes of Excess Female Mortality

I turn now to estimating the importance of different health investments in explaining this excess female mortality in childhood. I separate this discussion into three parts,

focusing first on the individual-level regression methodology, second on the direct calculation of medical effects and third on an analysis at the regional level.

1. Individual Level Regression

The individual-level regression methodology will only be useable when considering the effect of vaccinations, and only using surveys from the first NFHS. Information on malnutrition and medical treatment is not collected for children who have died, and later surveys did not ask about vaccinations for deceased children. Without this information it is not possible to construct the necessary individual-level panel.

Table 3 compares the results of estimating equations (2) and (3), which will evaluate the effect of vaccination differences on mortality differences. The regression is limited to children born between four and five years ago and the dependent variable is a dummy for having died between 18 months and four years, conditional on having lived to 18 months. The sample size is smaller than for the similar age group in Table 1 because I use only children born four to five years ago, not all children born in the last 10 years, and because I only use the data from the 1992 NFHS, since the 1998 NFHS did not ask about vaccinations for children who were dead. Column 1 estimates equation (2) and Column 2 estimates equation (3). As discussed in the methodology section, the share explained is calculated as the difference in the interaction coefficient divided by the interaction coefficient in Column 1. Vaccinations have a significant negative effect on mortality. Moving from zero vaccinations to complete vaccination decreases the probability of dying between ages one and four by a full 1.8%. In addition, vaccinations seem to explain a large share of the sex imbalance, around 30%. The standard errors are sufficiently large that I cannot reject equality of the coefficients (i.e. I cannot reject that the amount explained is equal to zero). However, the size of the point estimate is certainly economically significant.

One possible weakness of this analysis is recall bias. The regression includes controls for both vaccinations reported on the health card and vaccinations reported by the mother. If mothers in India are less likely to remember vaccinations for girls who have died, relative to boys who have died, there is a potential bias. This would have the effect of omitting a measure of true vaccination status, while including a measure of reported vaccination status. If true vaccination status (controlling for reported vaccination status) is correlated with the $Girl \times India$ interaction and with mortality, then the coefficient may be biased. This concern is ameliorated, at least

somewhat, by the inclusion of both measures of vaccination. Marks on the health card are likely to be a much better measure of actual vaccination status than maternal reports and the closer they get to the true vaccination, the less of a concern the omitted variable bias is. Further, including only the control for vaccinations reported on the health card makes relatively little difference in the results.

It is also possible that the effect of vaccinations vary by sex. If the benefit of vaccination for girls is larger than the benefit for boys then the share of the bias explained by vaccination may be understated. Although there is some evidence on sex differences in the nonspecific protective effect of vaccinations (Aaby et al, 2002), these do not seem to be consistent across vaccines. As a sensitive analysis, I repeat the regressions in Table 3, allowing for the effect of vaccination to differ by sex. The results (available from the author) are virtually identical.

2. Direct Calculation of Medical Effects

The second methodology here relies on direct evidence on the effect of child investments on mortality. In contrast to the regression framework, this analysis will be possible for all of the investments considered: malnutrition, treatment for diarrhea, treatment for respiratory infections and vaccinations. I consider only measles vaccination because this is the illness for which we have the best estimates of the effectiveness of vaccination. Obviously, the effect of measles alone will be an understatement of the total vaccine effect.

The calculations here require two elements: the difference in treatment by sex (D) and the effect of the treatment on mortality (μ). The first element is estimated in Table 4, which shows the sex bias in an indicator for severe malnutrition (Panel A), treatment for diarrhea (Panel B), treatment for respiratory infections (Panel C) and measles vaccination (Panel D). Controls are listed at the bottom of the table. The results indicate that boys in India are about 1 percentage point less likely to be malnourished and that this effect is significant. The results on medical treatment are mixed: boys are significantly more likely to be treated for respiratory infections, but not any more likely to be treated for diarrhea. The largest observed effects are for measles vaccination; boys are approximately 7 percentage points more likely to be vaccinated.

Information on the second element – the effect of treatment on mortality – is

presented in Appendix B. The details of the calculations appear in the Appendix but in general I use one of two techniques. In the case of malnutrition, I take advantage of studies in which nourishment levels of children were observed. The children were followed over time and mortality outcomes were reported. The difference in mortality by nutritional status provides an estimate of the effect of malnutrition. In the case of treatment and vaccination, the effect is the product of the probability of dying from the illness (either diarrhea, acute lower respiratory infections or measles) and the protective effect of treatment. In the case of measles, the effect of vaccination on mortality is the chance of dying from measles in India during this period multiplied by the effect of measles vaccination on measles mortality. The studies suggest that the protective effect of being well nourished is the largest, although the effect of measles vaccination is much larger than treatment for illnesses.[11] The studies used here are based on information from the developing world, or from India directly, so they should capture the experience of South Asia reasonably accurately.

Table 5 brings together the results from Table 4 and Appendix B and presents them with reference to the size of the sex imbalance. The first row of the table shows the excess female mortality between one and four years and the share explained is simply the sex difference multiplied by the mortality effect, divided by this baseline difference. The results here suggest that food plays a sizable role in the sex imbalance (explaining around 20%), but that treatment for diarrhea and respiratory infections plays only a limited one. The reason for this is straightforward. In the case of diarrhea, there is virtually no difference in treatment propensity. In the case of respiratory infections, there *is* a large difference in treatment propensity, but the chance of dying from that cause is not that large and the protective effect of treatment is small. The effect of the measles vaccine provides a supportive robustness check on the earlier estimates of the effectiveness of vaccination from the individual-level regressions. Measles vaccination alone explains about 21% of the sex imbalance. Although this is less than the 28% estimated in Table 3, it is an estimate for only one of many vaccinations.

[11]The larger effect of malnutrition does not seem to be an artifact of the difference in methodology. Using the DHS data from Africa, it is possible to get an estimate of the effect of measles vaccination on mortality which effectively parallels the estimate of malnutrition. The result suggests around a 3 percentage point decrease in death probability with measles vaccination, similar to what is seen in Appendix B. Although I do not use this estimate, because the goal is to use estimates from outside these data, it does provide some comfort.

3. Regional Level Analysis

The results in Tables 3 and 5 suggest that around 45% to 50% of the sex imbalance up to age five can be explained by vaccination, food intake and medical treatment. One issue, however, is that these variables may not be independent. If malnutrition makes children more likely to die from measles, the effect of malnutrition in Appendix B is also partially an effect of measles vaccination. This may lead the results in Table 5 to overstate the total explanatory power of the investments considered. Without an individual-level panel in which we observe all elements of food and treatment over time, this is a difficult problem to solve.

One option, however, is to collapse the data to the area level within India and then run regional-level equivalents to equations (2) and (3). By doing the same analysis specified for the individual regression, I can infer the share of mortality explained by different investments. There are clear issues with this approach. States within India differ on many dimensions and it may be difficult to fully control for these differences (I attempt to do so with controls for education, durable good ownership and parental preferences, but fully controlling will be virtually impossible). However, the advantage of the approach is that I can consider the effect of all investments simultaneously, which provides a useful robustness check.

The results of this analysis are in Table 6.[12] I show only the regression with no components of Z and the regression will all of the components of Z. What I will be able to conclude, therefore, is what share of the bias is explained by all of these elements together. The regression includes controls for average education level, average durable good ownership and average ideal sex ratio reported (these are also at the sex-region level). The data here are limited to 1992. The explanatory power is similar to what would be expected based on the other analyses. Around 40% of the sex imbalance is explained by these components together.

There are obvious limitations to this approach. Nevertheless, the results are roughly consistent with the previous ones. At least some significant share of the sex imbalance – perhaps close to half – seems to be explained by two factors: vaccination and food intake. Of course, this result implies that at least half of the imbalance remains unexplained. One possibility is that, with more accurate data, more of the

[12]This analysis is run using India only. The sample sizes for Africa are much smaller, allowing for only a very limited number of regions, making the comparison difficult. For simplicity, I assume that the

imbalance could be explained. Another possibility is that there are important elements not considered here – for example, direct parental intervention.

6. Conclusion

This paper uses a new methodology to analyze the proximate sources of excess female mortality in India. I argue that childhood is the most crucial time period in the sense that by the age of five there is enough excess female mortality to explain the entire sex imbalance in population in India. During this period, sex differences in both vaccination and nutrition play a large role in the excess mortality. I find roughly 50% of the sex imbalance remains unexplained by differences in vaccination, nutrition or medical care.

The first of these results stands somewhat in contrast to the situation in China, where high sex ratios at birth, or appearing immediately after, seem to drive high sex ratios. Das Gupta et al (2003) also note the apparent differences in these situation: the higher sex ratio in China seems to be driven by sex-selective abortion or sex-selective infanticide, while the situation in India, as demonstrated here, points more to childhood neglect. Understanding why these patterns differ is beyond the scope of this paper. However, it may be a fruitful direction for future work, especially as it suggests the growth of sex-selective technologies may have different impacts in different areas.

The second result here – the importance of vaccinations, in particular – may shed some light on the patterns seen in the first result. In particular, in the age breakdown we see that although mortality is quite high from ages 0 to 6 months, this time period does not seem to play a role in explaining excess female mortality. Initially, this contrast between the role of this time period in levels versus its role in differences may seem puzzling, but the role of vaccinations is consistent with this. Since vaccinations do not until a few months after birth (most between 6 and 12 months), we would expect differences due to vaccination to appear later in childhood, as we see in the data.

The results here have clear, and potentially important, policy implications. There has been significant focus in India on changing preferences – encouraging people to put greater value on women, promoting female schooling, etc. These are clearly useful goals. However, in the shorter run (before individual preferences can be

---

coefficient on "girl" in the regression should be zero, understanding that this is not exactly correct.

changed) the results here argue that, in particular, investments in vaccination for girls would have a direct effect on excess female mortality. Given a choice between the child investments discussed here, focus on universal (or, not sex-biased) vaccination would have the largest effect on mortality.

It is worth noting that in recent years the availability of sex-selective abortion has shifted some sex selection to the before-birth time period (Jha et al, 2006). If pre-birth sex selection is less costly than neglect then we would expect, in the limit, no post-birth treatment differences by sex. If parents can choose the sex of their child with certainty, then every girl who is born would be wanted, as would every boy, and we would not expect neglect after birth. It remains to be seen, however, just how large a phenomenon prenatal sex selection will become. In particular, the fact that gender differences in mortality show up not right at birth (i.e. from infanticide) but later in childhood (from neglect) may limit the eventual role for sex-selective abortion, which is a close substitute for infanticide, but not for neglect. Families may not have strong enough preferences to move to sex-selective abortion, but may still engage in less immediately obvious forms of discrimination, such as lack of vaccination. If this is true, then even in the world where sex-selection is available, the policy issues outlined here will be salient. Further, regardless of the long-run situation, it is clear in the short run that investments in health care for girls could save thousands of lives in India.

Appendix A: The Importance of the Childhood Sex Imbalance

The calculation driving Table 2 uses model life tables and stable populations from Coale et al (1983). I use the West Model life tables, with a mortality level of 17 and a gross reproduction rate of 2. This corresponds relatively closely to the experience of India in the mid 1990s. The first table below simply shows the mortality rates and stable population shares for this mortality level and GRR, from Coale et al (1983), for reference.

Age	Male Mort. Prob.	Female Mort. Prob	Sh. Men	Sh. Wom
0-1	8.68%	7.12%	2.88%	2.79%
1-5	3.53%	3.35%	10.45%	10.22%
5-10	1.19%	1.11%	11.80%	11.55%
10-15	0.89%	0.86%	10.67%	10.45%
15-20	1.39%	1.26%	9.63%	9.43%
20-25	1.96%	1.66%	8.65%	8.49%
25-30	2.10%	1.92%	7.73%	7.61%
30-35	2.39%	2.21%	6.90%	6.80%
35-40	2.91%	2.57%	6.13%	6.06%
40-45	3.77%	3.04%	5.41%	5.38%
45-50	5.04%	3.81%	4.72%	4.74%
50-55	7.07%	5.26%	4.05%	4.13%
55-60	9.97%	7.30%	3.39%	3.54%
60-65	14.46%	10.97%	2.72%	2.94%
65-70	20.70%	16.37%	2.06%	2.32%
70-75	29.82%	25.13%	1.42%	1.90%
75-80	42.36%	37.34%	0.83%	1.07%
80+			0.41%	0.79%

I first calculate the expected sex ratio (women divided by men) in the population using the observed sex ratio at birth. To do this, I assume that at birth there are 1000 women for 1069 men and calculate the expected sex ratio at the end of each age group based on the mortality patterns by age. I then weight the sex ratios by the share in that age group, which results in the total sex ratio. This calculation is shown in the table below.

Age	# Men	# Women	Sex Ratio (F/M)
0	1069.000	1000.000	0.935
1	976.232	928.840	0.951
5	941.761	897.715	0.953
10	930.573	887.777	0.954
15	922.319	880.124	0.954
20	909.527	869.052	0.955
25	891.682	854.600	0.958
30	872.947	838.183	0.960
35	852.067	819.684	0.962
40	827.288	798.659	0.965
45	796.108	774.364	0.973

50	756.008	744.884	0.985
55	702.566	705.733	1.005
60	632.541	654.243	1.034
65	541.095	582.485	1.076
70	429.088	487.109	1.135
75	301.134	364.694	1.211
80	173.574	228.528	1.317
Sex Ratio Weighted by Age Shares			**0.969**

The observed sex ratio in the population is around 0.932, quite different from the 0.969 predicted based on normal mortality patterns and the observed sex ratio at birth. The second step is to consider the same calculation, while assuming that the sex ratios up to age 5 are equal to those actually observed in the Indian population. This calculation appears in the next table.

Age	**# Men**	**# Women**	**Sex Ratio (M/F)**
0	1069.000	1000.000	0.935
1	1071.000	1000.000	0.934
5	1094.000	1000.000	0.914
10	1081.003	988.930	0.915
15	1071.415	980.405	0.915
20	1056.554	968.072	0.916
25	1035.825	951.973	0.919
30	1014.062	933.685	0.921
35	989.806	913.079	0.922
40	961.022	889.659	0.926
45	924.801	862.595	0.933
50	878.219	829.756	0.945
55	816.138	786.144	0.963
60	734.793	728.787	0.992
65	628.564	648.854	1.032
70	498.451	542.610	1.089
75	349.813	406.247	1.161
80	201.632	254.567	1.263
Sex Ratio Weighted by Age Shares			**0.933**

The predicted sex ratio is now 0.932, extremely close to the empirically observed ratio of 0.933. This suggests that the observed population sex ratio is consistent with a standard West Life Table mortality path *if* we assume that the sex ratio at age 5 is equal to what is observed in the Indian population. In other words, nearly all of the imbalance in the population sex ratio can be explained by events that occur before the age of 5.

Appendix B: Estimating the Effect of Medical Treatment on Mortality

This table shows estimates of the effect of medical treatment, or nutrition, on mortality drawn from existing literature. In each case, an effort is made to estimate the difference in probability of death for children that do, or do not, receive some beneficial treatment. For nutrition, the "treatment" is not being severely malnourished. For diarrhea, the treatment is oral rehydration solution. For respiratory infection, the treatment is doctor availability in general. Finally, for measles vaccination the treatment is, obviously, measles vaccination.

In all cases an effort is made to estimate the change in the total probability of death between the ages of 6 months and 5 years. The details of the calculations are below.

Study	**Effect**	**Methodology**
		Panel A: Food
Chen et al, 1980	16.44%	Children were weighed, measured and followed for two years. Mortality rate is calculated based on an extension of the two-year window to the entire period.
Sommer and Lowenstein, 1975	20.20%	Children were weighed, measured and followed for 18 months; this study is divided by age group. For each age group (ages 1, 2, 3 and 4 at the start), I use the data to calculate the mortality rate within one year of study enrollment. Denote these probabilities of death as q_1, q_2, q_3, q_4. The overall chance of mortality between ages of 1 and 5, conditional on having lived to the age of 1, is then $1-(1-q_1)(1-q_2)(1-q_3)(1-q_4)$. This product gives the total four-year probability of death from malnutrition.
Panel B: Diarrhea		
Rahaman et al, 1979	1.77%	The study reported the chance of dying in a village with 80% treatment and the chance of dying in a village with 40% treatment. I used this information to interpolate the chance of dying with no treatment and the chance of dying with treatment, and calculate a protective effect. I then calculate the share of deaths attributed to diarrhea in this area from Murray et al (1996) and multiply that number by the chance of dying in this age group (assumed to be 8%), which gives the total chance of dying from diarrhea. I assume this is the chance of dying if not treated, calculate the chance of dying if treated, and subtract.
Panel C: Respiratory Infections		
Ali et al, 2001	1.03%	The study reports the chance of dying in places with acute lower respiratory infection (ALRI) treatment and non-treatment programs. I used this information to calculate a protective effect. I then calculate the share of deaths attributed to ALRI in this area from Murray et al (1996) and multiply that by the chance of dying in this age group (assumed to be 8%), which gives the

		total chance of dying from ALRI. I assume this is the chance of dying if not treated and calculate the chance of dying if treated, and subtract.
Fauveau et al, 1992	0.98%	Same methodology as above.
Panel D: Measles Vaccination		
Clemens et al, 1988	2.88%	Case control study of measles vaccination in Bangladesh. Compared to the matched controls, vaccinated children had 36% lower mortality. I combine this with the baseline probability of dying after the first six months but before age five in regions with little or no vaccination (assumed to be 8%). I then multiply these two calculations for the result.
Koenig et al, 1990	3.68%	Same methodology as above; in this case, vaccinated children had a 46% lower mortality.

References

[1] **Aaby, Peter, H enrikJensen, May-Lill Garly, Carlitos Balé, Cesario Martins, and Ida Lisse**, "Vaccinations and child survival in a war situation with high mortality: effect of gender," *Vaccine*, 2002, (1-2), 15–20.

[2] **Agnihotri, Satish**, *Sex Ratio Patterns in the Indian Population: A Fresh Exploration*, New Delhi, India: Sage, 2000.

[3] **Agnihotri, Satish, Richard Palmer-Jones and Ashok Parikh**, "Missing Women in Indian Districts: A Quantitative Analysis," *Structural Change and Economic Dynamics*, 2002, 13, 285-314.

[4] **Ali, Mohammad, Michael Emch, Fahmida Tofail, and Abdullah H. Baqui**, "Implications of Health Care Provision on Acute Lower Respiratory Infection Mortality in Bangladeshi Children," *Social Science and Medicine*, 2001, (2), 267–277.

[5] **Basu, Alaka**, "Is Discrimination in Food Really Necessary for Explaining Sex Differentials in Childhood Mortality?," *Population Studies*, 1989, (2), 193–210.

[6] **Borooah, Vani**, "Gender bias among children in India in their diet and immunization against disease," *Social Science and Medicine*, 2004, , 1719–1731.

[7] **Chen, Lincoln, Alauddin Chowdhury, and Sandra Huffman**, "Anthropometric Assessment of Energy-Protein Malnutrition and Subsequent Risk of Mortality Among Pre-School Aged Children," *American Journal of Clinical Nutrition*, 1980, , 1836–1845.

[8] **Clemens, John D., Bonita F. Stanton, J. Chakraborty, Shahriar Chowdhury, Malla R. Rao, Mohammed Ali, Susan Zimicki, and Bogdan Wojtyniak**, "Measles vaccination and childhood mortality in rural Bangladesh," *American Journal of Epidemiology*, 1988, (6), 1330–1339.

[9] **Coale, Ansley**, "Excess Female Mortality and the Balance of the Sexes: An Estimate of the Number of 'Missing Females'," *Population and Development Review*, 1991, 517–523.

[10] **Coale, Ansley, Paul Demeny and Barbara Vaughan.** *Regional Model Life Tables and Stable Population,* Princeton: Princeton University Press, 1983.

[11] **Das Gupta, Monica**, "Selective Discrimination Against Female Children in Rural Punjab, India," *Population and Development Review*, 1987, (1), 77–100.

[12] **Fauveau, Vincent, M.K. Stewart, J. Chakraborty, and S.A. Khan**, "Impact on mortality of a community-based programme to control acute lower respiratory tract infections," *Bulletin of the World Health Organization*, 1992, (1), 109–116.

[13] **Griffiths, Paula, Zoe Matthews, and Andrew Hinde**, "Gender, family, and the nutritional status of children in three culturally contrasting states of India," *Social Science and Medicine*, 2002, , 775–790.

[14] **Jha, Prabhat, Rajesh Kumar, Priya Vasa, Neeraj Dhingra, Deva Thiruchelvam, and Rahim Moineddin**, "Low male-to-female sex ratio of children born in India: National Survey of 1.1 Million Households," *Lancet*, 2006, , 211–218.

[15] **Kishor, Sunita**, "'May God Give Sons to All': Gender and Child Mortality in India," *American Sociological Review*, 1993, (2), 247–265.

[16] **Klasen, Stephan and Claudia Wink**, "A Turning Point in Gender Bias in Mortality? An Update on the Number of Missing Women," *Population and Development Review*, 2002, (2), 285–312.

[17] **Koenig, Michael Alan, Mehrab Ali Khan, Bogdan Wojtyniak, John D. Clemens, J Chakraborty, Vincent Fauveau, James F. Phillips, Jalaluddin Akbar, and Uday S. Barua**, "Impact of measles vaccination on childhood mortality in rural Bangladesh," *Bulletin of the World Health Organization*, 1990, (4), 441–447.

[18] **Mishra, Vinad, T.K. Roy, and Robert Retherford**, "Sex Differentials in Childhood Feeding, Health Care, and Nutritional Status in India," *Population and Development Review*, 2004, (2), 269–295.

[19] **Murray, Christopher and Alan Lopez**, *The global burden of disease : a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*, Cambridge, MA: Harvard School of Public Health, Harvard University Press, 1996.

[20] **Murthi, Anne-Catherine Guio Mamta and Jean Dreze**, "Mortality, Fertility and Gender Bias in India: A District Level Analysis," *Population and Development Review*, 1995, (4), 745–782.

[21] **Oster, Emily**, "Hepatitis B and the Case of the Missing Women," *Journal of Political Economy*, 2005, (6), 1163–1215.

[22] **Pande, Rohini**, "Selective gender differences in childhood nutrition and immunization in rural India: The role of siblings," *Demography*, 2003, (3), 395–418.

[23] **Qian, Nancy**, "Missing Women and the Price of Tea in China: The Effect of

Relative Female Income on Sex Imbalance," 2007. MIT Working Paper.

[24] **Rahaman, M.M., K.M. Aziz, Y. Patwari, and M.H. Munshi**, "Diarrhoeal mortality in two Bangladeshi villages with and without community-based oral rehydration therapy," *Lancet*, 1979, (2), 809–812.

[25] **Rosenzweig, Mark and T. Paul Schultz**, "Market Opportunities, Genetic Endowments and Intrafamily Resource Distribution: Child Survival in India," *American Economic Review*, 1982, (4), 803–815.

[26] **Sen, Amartya**, "Missing Women," *British Medical Journal*, 1992, , 587–588.

[27] **Sen, Amartya,** "More than 100 Million Women are Missing," *New York Review of Books*, December 20, 1990.

[28] **Sommer, Alfred and Matthew S. Loewenstein**, "Nutritional status and mortality: a prospective validation of the QUAC stick," *American Journal of Clinical Nutrition*, 1975, , 287–292.

[29] **Visaria, P**., "The Sex Ratio of the Population of India," *Monograph No. 10, Census of India*, 1971.

[30] **Wells, Jonathan**, "Natural Selection and Sex Differences in Morbidity and Mortality in Early Life," *Journal of Theoretical Biology*, 2000, , 65–76.

Table 1. ***Sex Imbalance in Death by Age, All of India***

Dependent Variable: Child Died in Given Age Range (0/1)							
	< 6mons	6mons-1yr	1-2yrs	2-4yrs	4-6yrs	6-8yrs	8-10yrs
Explanatory Variables:							
Girl	−.0104***	.0002	−.0011*	−.0014***	−.0000	−.0000	.0007
	(−6.67)	(.33)	(−1.81)	(−2.66)	(−.04)	(−.02)	(1.21)
Girl × India	.0016	.0021***	.0055	.0041***	.0026***	.0005	−.0006
	(.91)	(2.86)	(6.48)	(5.57	(3.23)	(.70)	(−.91)
India	−.0077***	−.0135***	−.0175***	−.0098***	−.0088***	−.0032***	−.0008
	(−5.00)	(−16.12)	(−18.61)	(−12.65)	(−10.29)	(−4.58)	(−1.16)
Wealth (Durables)	−.0049***	−.0012***	−.0019***	−.0020***	−.0029***	−.0015***	−.0012***
	(12.77)	(−6.96)	(−9.21)	(−11.14)	(−13.98)	(−8.19)	(−6.63)
Child Age	−.0002***	−.0004***	−.0000	−.0000	.0007***	.0003***	.0010***
	(−6.61)	(−3.89)	(.42)	(−.39)	(5.56)	(2.12)	(4.64)
Mother's Age	−.0026***	−.0004***	−.0005***	−.0002***	−.0003***	−.0001	−.0000
	(−27.69)	(−7.42)	(−8.90)	(−4.95)	(−6.28)	(−1.31)	(−.46)
Mother's Educ.	−.0014***	−.0004***	−.0007***	−.0003***	−.0003***	−.0002***	−.0000
	(−13.27)	(−7.46)	(−10.85)	(−6.43)	(−6.33)	(−4.61)	(−.24)
# Siblings	.0257***	.0039***	.0045***	.0028***	.0025***	.0008***	.0005***
	(67.13)	(19.33)	(20.05)	(15.42)	(12.69)	(5.26)	(3.53)
Year	−.0007***	−.0004***	−.0001	.0001	−.0001	.0000	.0000
	(−4.57)	(−5.97)	(−1.03)	(1.50)	(−1.18)	(.16)	(−.42)
Number of Observations	338,943	294,410	278,397	243,496	211,896	150,801	86,604

Notes: All regressions are limited to children born in the last ten years. The dependent variable is a dummy equal to one if the child died within the specified age range *conditional on* having lived up until that age category. The measure of wealth is a measure of the number of durable goods owned, where the durable goods are radio, television, refrigerator, motorcycle, bicycle and car. Controls for birth order (dummies) are also included in all regressions. India is observed in 1992 and 1998. The African countries included are Ethiopia (2000), Kenya (1998, 2003), Malawi (2000), Namibia (1992, 2000), Tanzania (1996, 1999) and Zambia (1996, 2001).

t-statistics in parentheses

***** significant at 10%; ****** significant at 5%;******* significant at 1%

–

Table 2. ***Share of Overall Imbalance Explained by Age 5***

	All India
Observed Sex Ratio (F/M) Age 5	0.914
Predicted Sex Ratio, Model Mortality after 5	0.933
Empirical Population Sex Ratio	0.932

Notes: Sex ratio is the number of women divided by the number of men. Observed sex ratio at age 5 is from the NFHS. The predicted population sex ratio is calculated by assuming that the sex ratio at age 5 is naturally occurring and then assuming that mortality is equal to what is reported in life tables after age 5. Mortality is taken from the Coale, Demeny and Vaughn (1983) life tables, using the West Model mortality level 17 with a GRR of 2.0. Empirical population sex ratio is from the 1991 Census.

Table 3. ***Impact of Vaccines on Excess Female Mortality***

Dependent Variable: Child Died 1 year - 4 years		
	(1)	(2)
Explanatory Variables:		
Girl	−.0013	−.0007
	(-.44)	(-.29)
Girl × India	.0091**	.0065
	(2.33)	(2.03)
India	−.0233***	−.0393
	(−5.58)	(−8.13)
# Vacc Reported by Mom		−.0009***
		(−3.72)
# Vacc on Health Card		−.0026***
		(−6.57)
Obs.	13,817	13,817
Share Explained	**28.4%**	

Notes: Controls in all regressions include dummies for child size at birth, maternal age, maternal education, child age, income (durables), total number of children and dummies for birth order. The two measures of vaccines capture both the total number of vaccines reported by the mother and the total number marked on the health card (3 DPT vaccines, 2 polio vaccines, measles and BCG). Data comes from the 1992-1993 NFHS
t-statistics in parentheses
***** significant at 10%; ****** significant at 5%;******* significant at 1%

Table 4. ***Sex Imbalance in Malnutrition and Treatment of Illness***

Panel A: Malnutrition	
Dependent Variable: Child is Severely Malnourished	
Explanatory Variables:	
Girl	−.0025
	(−1.17)
Girl × India	.0116***
	(4.39)
India	.0342***
	(18.86)
Number of Observations	80819
Panel B: Diarrhea	
Dependent Variable: Child received treatment if had Diarrhea	
Girl	−.0176
	(−1.46)
Girl × India	−.0027
	(−.16)
India	−.1057***
	(−8.64)
Number of Observations	13666
Panel C: Respiratory Infections	
Dependent Variable: Child received treatment if had Cough or Fever	
Girl	−.0015
	(−.21)
Girl × India	−.0357***
	(−3.24)
India	.0992***
	(12.55)
Number of Observations	26,077
Panel D: Measles Vaccination	
Dependent Variable: Child is Vaccinated for Measles	
Girl	.0356*
	(1.66)
Girl × India	−.0698***
	(−2.98)
India	−.4110***
	(−27.64)
Number of Observations	15,120

Notes: Controls in all regressions include income, age of child, dummies for birth order, number of children in the household, maternal education and age. Data is from the 1992-1993 NFHS and the 1998-1999 NFHS. Panel A, B and C: Children aged 1-4 years; Panel D: children aged 3-4 years.
t-statistics in parentheses
***** significant at 10%; ****** significant at 5%;******* significant at 1%

Table 5. ***Share of Missing Girls Explained by Food, Treatment and Vaccination***

	All India
Baseline Difference	0.0111
Food (% Explained)	19.1%
Diarrhea Treatment (% Explained)	0.43%
ALRI Treatment (% Explained)	3.9%
Measles Vaccination (% Explained)	20.6%

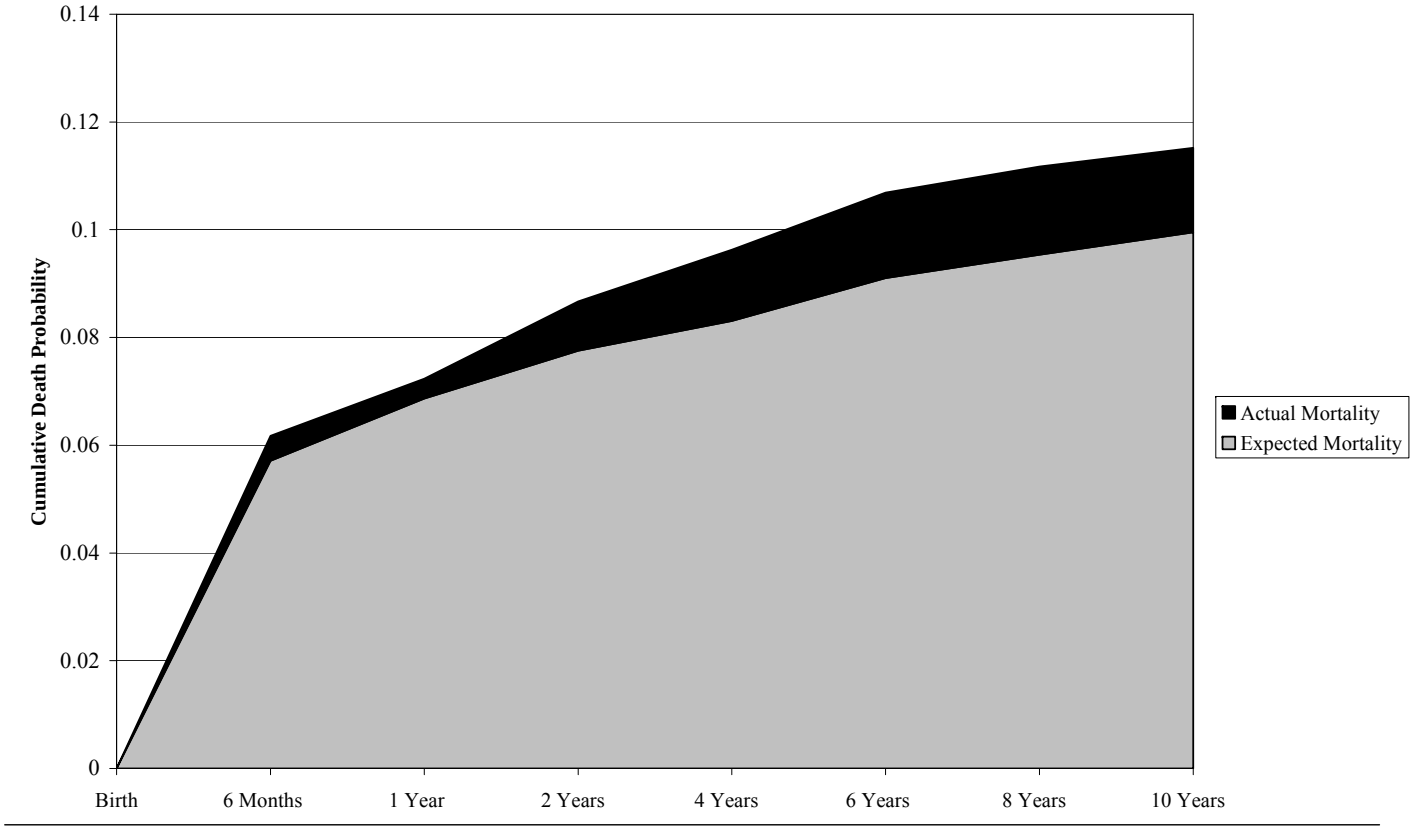
Notes: The baseline difference is the value of the coefficient on the variable "girl x India" in a regression of the form of Table 1, but aggregating death rates for children aged 6 months to 5 years. The share of the puzzle explained by each indicator is equal to the difference from Table 4 multiplied by the effect on mortality from Appendix B and then divided by the baseline difference. In the case of ALRI and Measles, the effect used from Appendix B is the average of the two study effects.

Table 6. ***Regional Analysis of Proximate Causes***

Dependent Variable: Share of Girls/Boys in Region Who Died Ages 1-5		
	Without Controls	**With Controls for Vaccination, Malnourishment**
Explanatory Variables:		
Girl	.0156***	.0091**
	(4.03)	(2.52)
Vaccination		.0003
		(.27)
% Severely Malnourished		.0273
		(.89)
Vacc x Malnour.		.0018
		(.19)
Average Mother Education	−.0015	−0018*
	(−1.22)	(−1.67)
Average Wealth (Durables)	−.0029	−.0032
	(−.68)	(−.83)
Average Ideal Sex Ratio	.0411***	.0113
	(3.73)	(1.07)
Constant	−.0297	.0178
	(−1.60	(1.01)
Number of Observations	339	338
Share explained	**41.7%**	

Notes: Each observation represents a sex-region and the dependent variable is the share of children born 5-10 years ago in that sex-region group who died between the ages of 1 and 4 years (conditional on reaching 1). The measure of vaccination represents the first principal component of indicators for having gotten six vaccines: measles, DPT(1,2,3) and Polio(1,2). The data is from the 1992-1993 NFHS only. "Ideal Sex Ratio" is the average ideal number of boys divided by the average ideal number of girls.
t-statistics in parentheses
***** significant at 10%; ****** significant at 5%;******* significant at 1%

Figure 1:
Actual and Expected Female Mortality in India



Notes: The sample is limited to children born in the ten years before the survey, and estimates are based on the regressions in Table 1. Expected mortality is the predicted mortality from the regression for girls in India *if* they did not have the negative effect of being a girl in India (i.e. if the value of the interaction term was zero). Actual mortatlity is actual predicted mortality.