TI 2007-067/3
Tinbergen Institute Discussion Paper

What explains the Rural-Urban Gap in Infant Mortality — Household or Community Characteristics?

Ellen Van de Poel[a]
Owen O. Donnell[b]
Eddy Van Doorslaer[ac]

[a] *Erasmus University Rotterdam;*
[b] *University of Macedonia, Greece;*
[c] *Tinbergen Institute.*

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

What explains the rural-urban gap in infant mortality — household or community characteristics?

Ellen Van de Poel[a1], Owen O'Donnell[b] and Eddy Van Doorslaer[ac] *

[a] *Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands*

[b] *University of Macedonia, Greece*

[c] *Tinbergen Institute, The Netherlands*

[1] Correspondence to:
Ellen Van de Poel
Department of Applied Economics, Erasmus University Rotterdam
Burg. Oudlaan 50/P.O. Box 1738
3000 DR Rotterdam, The Netherlands
Tel: +31 10 408 1502,
Fax: +31 10 408 91 41
E-mail: vandepoel@few.eur.nl

* The authors are grateful to the Institute for Housing and Urban Development Studies for funding the project on "Urbanization, Health and Health Inequality", from which this paper derives.

ABSTRACT

The rural-urban gap in infant mortality rates is explained using a new decomposition method that permits identification of the contribution of unobserved heterogeneity at the household and the community level. Using Demographic and Health Survey data for six Francophone countries in Western Sub-Saharan Africa, we find that differences in the distributions of factors that determine mortality – not differences in their effects – explain almost the entire gap. Higher infant mortality rates in rural areas mainly derive from the rural disadvantage in household level characteristics; both observed and unobserved, which explain three-quarters of the gap. Among the observed characteristics, household environmental factors—potable water, electricity and quality of housing materials—are the most important contributors explaining 38% of the gap. Unobserved household level determinants explain 10% of the gap. Community level determinants explain 13% of the gap, including 3% that is due to unobservable community level heterogeneity.

KEYWORDS

Sub-Saharan Africa, rural-urban inequality, infant mortality, decomposition, unobserved heterogeneity

INTRODUCTION

Rural children face higher mortality rates than their urban counterparts (Cleland, Bicego & Fegan 1992; Knobel, Yang & Ho 1994; Brockerhoff 1995; Lalou & LeGrand 1997; Sastry 1997a; Gould 1998; Heaton & Forste 2003; Wang 2003; Cai & Chongsuvivatwong 2006). While the rural disadvantage in average child survival in developing countries is firmly established, its explanation is less clear. This paper seeks to redress the paucity of information on the causes of the rural-urban gap in infant mortality rates by using a new decomposition method that permits quantification of the contribution of unobserved heterogeneity at the household and the community level. Because of the limited availability of community level data, few studies of child survival have been able to focus on the relative roles of community and household characteristics (Sastry 1996). The distinction is nonetheless important since it is helps determine the most appropriate level for policy intervention. This paper exploits community level data on health facilities and public infrastructure but also identifies the contribution of unobservable community level characteristics. The decomposition is applied to data from six Francophone countries in West Sub-Saharan Africa, a region that is relatively understudied despite having infant mortality rates that are amongst the highest in the world (World Bank 2006).

Household level factors appear to be important in explaining rural-urban differences in child mortality. Van de Poel, O'Donnell & van Doorslaer (2007) found that controlling for differences in household wealth reduces the median rural-urban risk ratio in under-five mortality in a set of 47 developing countries by 59 percent. After controlling for a broad range of household socioeconomic and demographic factors, the urban advantage in child mortality remains significant in about one third of the countries. However, this study does not exploit any information on community characteristics, such as availability of health care services, which are integral to the differential conditions experienced in urban and rural locations and are potentially important contributors to the rural-urban disparity in infant mortality. Sastry (1996, 1997a) highlights the importance of community level factors in explaining the rural-urban infant mortality differential in Brazil. Lalou and LeGrand (1997) and Heaton & Forste (2003) provide evidence suggesting that the limited availability of health care is partly responsible for the lower survival chances of children born in the rural Sahel and rural Bolivia respectively. The present paper uses Demographic and Health Survey (DHS) data for Sub-Saharan African countries for which the latest round also had a community survey providing information on the availability of health care services and other community infrastructure. We explicitly distinguish between characteristics that vary at the community and household levels and further categorize the latter into proximate and socioeconomic determinants of child mortality (Mosley & Chen 1984). Besides these observed determinants of child survival, there are many household and community factors that might affect infant mortality but are not measured in the data. At the household level, these include biological and genetic factors, as well as cross-infection rates and health related behavior. At the community level, infant mortality might be influenced by specific cultures and customs, by geographical aspects such as climate and soil fertility and by the quantity and quality of infrastructure. To take account of these unobservable determinants of infant mortality at both the household and the community level, we use a three-level random intercept logistic regression model (Gibbon & Hedeker 1997; Sastry 1997b; Bolstad &

Manda 2001). Thereafter we explain the rural-urban gap in infant mortality by applying an Oaxaca-type decomposition for non-linear models as suggested by Fairlie (2005), that we extend to take account of the unobserved household and community level heterogeneity.

Data are from six Sub-Saharan African countries (Benin, Central African Republic, Chad, Guinea, Mali and Niger). With an average of 96 out of 1000 children dying before the age of one, Sub-Saharan Africa has the highest infant mortality burden in the developing world (World Bank 2006). Within this region, infant mortality levels are among the highest in West (mostly Francophone) Africa (excluding Ghana) (Kuate-Defo & Diallo 2002). However, most of the published health research based on this region has focused on Anglophone countries. Attention to reproductive health in Francophone Africa developed much later than in other regions. For many years after independence, most of the countries operated under pronatalist policies. Family planning services were not introduced into national health programs until the mid- to late 1980s, which was due in part to a 1920 French law forbidding abortion and promotion of contraceptives. The law has now been repealed in all of the countries studied except Benin and Mali, and in these two cases it is no longer enforced. Population policies have evolved in all of the countries, albeit at varying speeds (Tantchou & Wilson 2000). Rural-urban differences in infant mortality rates are marked in the region. On average across the six countries studied, mortality in rural areas exceeds that in urban areas by five deaths per 100 births. If infant mortality rates in rural areas were reduced to those in urban areas, about 80,000 fewer children would die each year in these countries.[1]

In the remainder of the paper we first describe the data on infant mortality and categorize the variables used to explain its variation according to a coherent conceptual framework. This is followed by presentation of the methodology used to model infant mortality allowing for unobservable heterogeneity at the household and community levels and to decompose its difference across rural and urban locations. Thereafter results are presented; we first focus on the rural-urban gap in infant mortality rates when pooling the data across all countries and then consider country specific results. We find that differences in the distributions of factors that determine mortality – not differences in their effects – explain almost the entire gap. Higher infant mortality rates in rural areas mainly derive from the rural disadvantage in household level characteristics; both observed and unobserved, which, in the pooled analysis, explain three-quarters of the gap. Among the observed characteristics, household environmental factors, such as supply of potable water and electricity, and the quality of housing materials, are the most important contributors explaining 38% of the gap. Unobserved household level determinants explain 10% of the gap. Community level determinants explain 13% of the gap, including 3% that is due to unobservable community level heterogeneity. The final section concludes with a summary of the results and an interpretation of their implications.

DATA

Infant mortality

The most recent round of the Demographic Health Surveys (DHS) of Sub-Saharan African countries includes a survey on community characteristics in seven countries: Benin, Central African Republic (CAR), Chad, Gabon, Guinea, Mali and Niger. Table 1 shows the survey years, sample sizes and estimated infant mortality rates, expressed as the proportion of all live-born children that die before reaching the age of one. Children born between 10 and 1 years before the survey are included in the sample. Figure 1 gives an idea of the geographical clustering on infant mortality in Sub-Saharan Africa. Even relative to this region, infant mortality rates are high in the countries included in this study, with the exception of Gabon, which is more prosperous than most nearby countries.[2]

Table 1 also shows infant mortality rates stratified by rural-urban location and the proportion of the population in urban settings. Except in Gabon, the larger part of the population is located in rural areas and suffers from significantly higher infant mortality than the urban population. The rural-urban gap is by far the largest in Niger. We do not include Gabon in the analysis presented hereafter since the rural-urban gap in infant mortality is insignificant and the country is quite distinct from the others.

The distributions of children and infant deaths per household for the data pooled across all countries are summarized in Table2. Very large households are more prevalent in rural locations; the median number of children per household in urban and rural areas is three and four respectively. There is clear evidence of clustering of deaths within households. In urban areas, 4% of households with more than one death account for 40% of all deaths. In rural areas, 48% of deaths are concentrated in the 7% of households with more than one death. In both rural and urban areas about 1% of the households contribute three or more deaths. Together, these families account for 14% of the total number of deaths in urban areas and 19% in rural areas. This clustering suggests that it may be important to allow for household level heterogeneity in modeling infant mortality.

The primary sampling unit (PSU) in the DHS is the community. Generally a rural community spans one village or settlement, whereas an urban community is a part of a city. On average about 20-25 women are interviewed within every urban PSU, and about 30-40 women within every rural PSU. If there are more than 400 households within a PSU, only a smaller segment is interviewed.

Table 3 shows the distributions of children and infant deaths per community for the data pooled across all countries. The average number of children per community is considerably larger within rural than within urban areas (57 versus 33 respectively), reflecting both the larger number of women interviewed within rural PSUs and the slightly larger household sizes in rural areas. Within urban areas, 7% of deaths occur in the 1% of communities with 10 or more deaths; whereas within rural areas 49% of deaths are concentrated in the 22% of communities with 10 or more deaths. These numbers suggest that clustering of infant deaths within communities is less pronounced than within households, and is even less so within urban areas. The latter seems to be partly caused by the much smaller number of deaths within urban areas.

Explanatory variables

Our conceptual framework for modeling infant mortality derives from Mosley & Chen (1984), who distinguish between proximate and socioeconomic determinants. The former are mostly biological risk factors with a direct aetiological impact on child mortality. Of the five categories of proximate determinants identified by Mosley & Chen, the DHS data only provide direct measures of what they refer to as ‘maternal factors’. We include mother’s age at birth, birth order and an indicator of short birth interval (<24 months). The effect of birth order is captured by a dummy for first born children and another for children with a birth order higher than four (Sastry 1997a; Rutstein 2000).[3] The importance of these variables has been confirmed quite extensively in the literature (see e.g. Curtis, Diamond & McDonald 1993; Ronsmans 1996; Sastry 1997a; Manda 1999; Folasada 2000; Bhargava 2003).

In the Mosley & Chen (1984) framework, socioeconomic factors impact on child health and survival through the proximate determinants. In the absence of data that perfectly captures all proximate determinants, socioeconomic factors should explain some of the residual variation in child survival.

Mosley & Chen distinguish between socioeconomic determinants at the individual, household and community levels. At the individual level, we include mother's education, which has been considered an important determinant of child mortality since the work of Caldwell (1979) and has subsequently been reaffirmed (see e.g. Cleland & van Ginneken 1988; Hobcraft 1993; Kalipeni 1993). We use a dummy variable indicating the mother has no formal education. Education may affect child survival chances through knowledge of health production but also through the empowerment of women within the household and the consequent priority given to child health in household resource allocation. We further control for the social status of women through the mother's age at first marriage (Folasada 2000; Bhargava 2003), the sex of the household head (Lloyd & Blanc 1996; Canagarajah 2001) and the mother's use of contraception (Birdsall & Chester 1987). Traditions, social norms and attitudes that may impact on investments in child health are also proxied by the age of the head of household and the sex of the child. Molbak *et al* (1997) found that children in households with a younger head are associated with higher diarrhea prevalence, a major contributor to infant mortality. Traditions, such as the payment of a dowry, may result in differential investments in the health of boys and girls (Rosenzweig & Schultz 1982; Tambiah *et al* 1989).

Household income and wealth can be expected to impact on child survival through the purchase of goods and services that limit exposure to risk factors or treat the consequences of such exposure. We separately include aspects of household wealth that may impact on child health through exposure to environmental contamination, which Mosley & Chen identify as one of the five proximate determinants of mortality. Hence, we include a dummy for household access to safe drinking water and another for satisfactory sanitation (see Table 4 for definitions) (Victora *et al* 2005). The health effects of such environmental health determinants were highlighted in the WHO's 2002 World Health Report (WHO 2002), which showed that unsafe water, poor sanitation, and hygiene are the cause of 4–8% of the overall burden of diseases in developing countries and nine-tenths of diarrheal diseases. Also Esrey *et al* (1991) and Hertz**,** Herbert & Landon (1994) have illustrated the strong associations between sanitation and child survival. Further, we include information on the floor material of the household dwelling and whether the household has an electricity supply (Smith, Ruel & Ndiaye 2005). Housing materials may act as a proxy for the quality of housing, exposure to vermin and overcrowding, which raises the risk of respiratory disease.[4] Electricity facilitates more hygienic preparation of food and sterilization.

Besides limiting exposure to environmental contamination, wealth can raise survival chances through the purchase of food, medicines and access to health care. To obtain a proxy for wealth, beyond that indicated by access to drinking water, sanitation, electricity and housing materials, we construct an index using principal components analysis on possession of assets such as a car, motor, bicycle, radio, television, and fridge (Filmer & Pritchett 2001; Hong 2006). The first principal component is used to divide households into the poorest, middle and richest thirds.[5]

At the community level, we approximate the availability of health care services and public transport with dummies to indicate the presence of a health facility and any public transport respectively.[6] Brenneman (2002) found evidence reported in various studies that better transport contributes to easier access to health care as well as easier staffing and operation of clinics. Moreover, improved transport policy can reduce air pollution in urban areas and increase the supply of food in rural ones.

Table 4 provides a description of all the explanatory variables included in the analysis categorized according to the Mosley & Chen framework. Table 5 shows summary statistics of all covariates across urban and rural areas. Children born in rural areas are at a disadvantage across virtually all health determinants. This is true for the pooled cross-country sample, as well as within each country.[7]

While the data allows us to measure many of the important determinants of infant mortality, one might expect there to be considerable variation in survival chances across households and communities that is not captured by these covariates. There is some indication of the presence of such household and community level heterogeneity from the clustering of deaths described in the previous sub-section. In the next section we present a model of infant mortality that allows for household and community level effects and then show how the contribution of these effects to the rural-urban gap, as well as those of the observable factors, can be quantified.

METHODS

Three-level random intercept logistic regression

We model the probability of infant mortality using a three-level logit with random intercepts representing unobservable heterogeneity at both the household and the community level (Gibbon & Hedeker 1997). Compared to standard logistic regression, this model has the advantage of estimating the correlation in survival probabilities among children belonging to the same family and that among those residing in the same community that persists after controlling for observed characteristics (Sastry 1997b; Bolstad & Manda 2001). Failure to account for this unobserved heterogeneity would lead to inconsistent coefficients.[8] The three-level random component logistic model can be written as:

$$\begin{aligned} y_{ihc}^{*} &= x_{ihc}\beta + \eta_{hc} + \eta_{c} + \varepsilon_{ihc} \\ y_{ihc} &= 1 \text{ if } \mathrm{y}_{\mathrm{ihc}}^{*} > 0 \end{aligned} \qquad (1)$$

where y_{ihc}^{*} is a latent index the sign of which determines observation of an infant death $\left(y_{ihc} = 1\right)$, the indices *i, h* and *c* refer to infants, households and communities respectively, and η_{hc} and η_{c} are the random household and community level intercepts respectively. The idiosyncratic error term $\left(\varepsilon_{ihc}\right)$ is assumed to follow a logistic distribution and the random intercepts at each level are assumed multivariate normal. The random intercepts at different levels are assumed mutually independent and also independent of the idiosyncratic error and the covariates, x_{ihc}. The latter exogeneity assumption could be challenged for a few of the regressors. In particular, high birth order and short birth interval would be endogenous if they reflect previous infant deaths resulting from the same unobservable factors that condition the survival chances of all children in a household (Bhargava 2003). It is difficult to allow for such endogeneity in the context of a three-level non-linear model. We have confirmed that the effects of the other regressors are robust to dropping the potentially endogenous ones and relying on the household intercepts to capture their effect.

The likelihood of the model can be written as:

$$\prod_{c=1}^{n}\int\left\{\prod_{h=1}^{n_c}\int f\left(y_{ihc}\mid x_{ihc},\eta_{hc},\eta_c\right)\phi\left(\eta_{hc}\mid x_{ihc}\right)d\eta_{hc}\right\}\phi\left(\eta_c\mid x_{ihc}\right)d\eta_c \qquad (2)$$

where $f(y_{ihc}\left|x_{ihc,}\eta_{hc},\eta_c\right.)=\prod_{i=1}^{n_{hc}}\Lambda\left(x_{ihc}\beta+\eta_{hc}+\eta_c\right)^{y_{ihc}}\left(1-\Lambda\left(x_{ihc}\beta+\eta_{hc}+\eta_c\right)\right)^{1-y_{ihc}}$ is the joint density of the dependent variable for all infants within a given household conditional on the household and community effects as well as the observable explanatory variables, $\Lambda(\ )$ is the logistic cumulative density function, $\phi()$ represents the normal densities of the of the random intercepts with variances standardized to unity , η indicates the number of communities, η_c denotes the number of households within any given community and η_{hc} the number of infants within a given household.

The 'posterior' (conditional) density function of the random effects can be calculated using Bayes' Theorem. For the household effects, this gives:

$$p(\eta_{hc}\left|y_{ihc},x_{ihc},\eta_c\right.)=\frac{f\left(y_{ihc}\mid x_{ihc},\eta_{hc},\eta_c\right)\phi\left(\eta_{hc}\right)}{f\left(y_{ihc}\mid x_{ihc},\eta_c\right)}=\frac{f\left(y_{ihc}\mid x_{ihc},\eta_{hc},\eta_c\right)\phi\left(\eta_{hc}\right)}{\int f\left(y_{ihc}\mid x_{ihc},\eta_{hc},\eta_c\right)\phi\left(\eta_{hc}\right)d\eta_{hc}} \qquad (3)$$

where $p(\ )$ denotes the posterior density. Because of the assumed independence between the household and community random effects and of each with the covariates, $\phi(\eta_{hc}\left|x_{ihc},\eta_c\right.)=\phi\left(\eta_{hc}\right)$ and the marginal distribution of η_c appears in both the numerator and denominator and so cancels out. Following from this, the posterior means of the random household effects are given by:

$$\hat{\eta}_{hc} = \frac{\int \eta_{hc} f\left(y_{ihc} \mid x_{ihc}, \eta_{hc}, \eta_c\right) \phi\left(\eta_{hc}\right) d\eta_{hc}}{\int f\left(y_{ihc} \mid x_{ihc}, \eta_{hc}, \eta_c\right) \phi\left(\eta_{hc}\right) d\eta_{hc}} \tag{4}$$

Similarly the posterior means of the community effects are given by:

$$\hat{\eta}_c = \frac{\int \eta_c \left\{ \prod_{h=1}^{n_c} \int f\left(y_{ihc} \mid x_{ihc}, \eta_{hc}, \eta_c\right) \phi\left(\eta_{hc}\right) d\eta_{hc} \right\} \phi\left(\eta_c\right) d\eta_c}{\int \left\{ \prod_{h=1}^{n_c} \int f\left(y_{ihc} \mid x_{ihc}, \eta_{hc}, \eta_c\right) \phi\left(\eta_{hc}\right) d\eta_{hc} \right\} \phi\left(\eta_c\right) d\eta_c} \tag{5}$$

The likelihood is maximized and the posteriors means of the random effects are computed by adaptive quadrature using the *GLLAMM* program in Stata (Rabe-Hesketh, Skrondal & Pickles 2002; Rabe-Hesketh, Skrondal & Pickles 2005).

Decomposition

Rural-urban disparity in infant mortality rates can arise from differences in: a) the distributions of observable determinants of infant mortality; b) the effects of those determinants; c) the distributions of unobservable determinants. Blinder-Oaxaca type decomposition can be used to quantify the relative importance of these three explanations (Blinder 1973; Oaxaca 1973). In standard decomposition, the difference in the mean effects of unobservables is reflected in the difference in the intercepts of urban and rural specific regressions. But these intercept differences are not particularly helpful in pinpointing the source of rural-urban disparities in infant mortality since they provide no information on the level at which unobservables operate. We provide a more detailed explanation of the rural-urban disparity by quantifying the contribution of unobservable determinants of infant mortality at both the household and community levels. This is achieved by extending the non-linear decomposition of the group difference in a binary indicator proposed by Fairlie (2005) to a three-level random intercept logistic regression model.[9]

The rural-urban gap in average infant mortality can be decomposed as follows:

$$\bar{Y}^r - \bar{Y}^u = \left[\sum_{i=1}^{N^r} \frac{\Lambda(x_{ihc}^r \hat{\beta} + \hat{\eta}_{hc}^r + \hat{\eta}_c^r)}{N^r} - \sum_{i=1}^{N^u} \frac{\Lambda(x_{ihc}^u \hat{\beta} + \hat{\eta}_{hc}^u + \hat{\eta}_c^u)}{N^u} \right] + \left[\sum_{i=1}^{N^u} \frac{\Lambda(x_{ihc}^u \hat{\beta}^r + \hat{\eta}_{hc}^u + \hat{\eta}_c^u)}{N^u} - \sum_{i=1}^{N^u} \frac{\Lambda(x_{ihc}^u \hat{\beta}^u + \hat{\eta}_{hc}^u + \hat{\eta}_c^u)}{N^u} \right] \quad (6)$$

where superscripts r and u indicate values of covariates/estimates obtained from the rural and urban samples of children respectively, N^r and N^u indicate the number of infants located in rural and urban areas respectively, $\hat{\beta}$ refer to the coefficients from the pooled (urban and rural) model and $\hat{\eta}_j^k$ $j = hc, c$ and $k = r, u$ are the household and community specific posterior means of the random intercepts that are estimated from (4) and (5). The term in the first set of brackets represents the part of the rural-urban gap that is due to differences in the distributions of the observable determinants of infant mortality as well as the differences in the unobservable household and community level determinants. The term in the second brackets gives the gap due to differences in the effects of the observable determinants.[10] The coefficients from the pooled (urban and rural) model are used to weight the differences in the x's in the first term, and the urban distribution of x's is used to weight differences in the coefficients in the second term.[11]

The gap can then be decomposed further into the contributions of each covariate, both through its distribution and its effect. However, we will focus on the contributions of differences in the distributions of covariates and random household and community effects since, as will become apparent below, differences in coefficients contribute only marginally to explanation of the rural-urban gap in infant mortality. To illustrate how the contributions of differences in the distributions of particular covariates are identified, consider a simple case in which infant mortality is explained by two determinants, x_1 and x_2, and $N^r = N^u$. The contribution of the difference in the distributions of x_1 to the rural-urban gap is then equal to (Fairlie, 2005)

$$\frac{1}{N^r}\sum_{i=1}^{N^r}\Lambda(\alpha + x_{1ihc}^r\widehat{\beta}_1 + x_{2ihc}^r\widehat{\beta}_2 + \widehat{\eta}_{hc}^r + \widehat{\eta}_c^r) - \Lambda(\alpha + x_{1ihc}^u\widehat{\beta}_1 + x_{2ihc}^r\widehat{\beta}_2 + \widehat{\eta}_{hc}^r + \widehat{\eta}_c^r) \qquad (7)$$

Similarly, the contribution of x_2 can be expressed as

$$\frac{1}{N^r}\sum_{i=1}^{N^r}\Lambda(\alpha + x_{1ihc}^u\widehat{\beta}_1 + x_{2ihc}^r\widehat{\beta}_2 + \widehat{\eta}_{hc}^r + \widehat{\eta}_c^r) - \Lambda(\alpha + x_{1ihc}^u\widehat{\beta}_1 + x_{2ihc}^u\widehat{\beta}_2 + \widehat{\eta}_{hc}^r + \widehat{\eta}_c^r) \qquad (8)$$

The contribution of the difference between rural and urban areas in the means of the unobservable household-level determinants can be estimated by[12]:

$$\frac{1}{N^r}\sum_{i=1}^{N^r}\Lambda(\alpha + x_{1ihc}^u\widehat{\beta}_1 + x_{2ihc}^u\widehat{\beta}_2 + \widehat{\eta}_{hc}^r + \widehat{\eta}_c^r) - \Lambda(\alpha + x_{1ihc}^u\widehat{\beta}_1 + x_{2ihc}^u\widehat{\beta}_2 + \widehat{\eta}_{hc}^u + \widehat{\eta}_c^r) \qquad (9)$$

Finally, the contribution of the difference in community-level heterogeneity is estimated by:

$$\frac{1}{N^r}\sum_{i=1}^{N^r}\Lambda(\alpha + x_{1ihc}^u\widehat{\beta}_1 + x_{2ihc}^u\widehat{\beta}_2 + \widehat{\eta}_{hc}^u + \widehat{\eta}_c^r) - \Lambda(\alpha + x_{1ihc}^u\widehat{\beta}_1 + x_{2ihc}^u\widehat{\beta}_2 + \widehat{\eta}_{hc}^u + \widehat{\eta}_c^u) \qquad (10)$$

Basically, the contribution of each variable to the gap equals the change in the average predicted probability of dying from replacing the rural distribution with the urban distribution of that variable while holding the distributions of the other variables constant.[13] Since in our case the urban sample is smaller than the rural, a random rural subsample is drawn and matched with the urban sample on the basis of predicted probabilities of dying (Fairlie 2005).[14] Since the results depend on the specific subsample that is drawn, the process is repeated 200 times and average results are reported.[15]

Regression estimates, as well as the random drawing of the rural subsample take into account the sample weights that come with DHS data. Weights are adjusted for differences in population size of countries (estimates obtained from the World Bank (2006)). This way, countries with large population size have relatively more influence on the pooled results and these can be interpreted as being representative for the region.[16]

RESULTS

Regression results

We first present results from the pooled cross-country analysis.

Table 6 shows regression coefficients for the total sample, and for urban and rural subsamples. Since the dependent variable indicates whether the child died within its first year, a positive coefficient means an increased risk of death. All coefficients have intuitive signs. We find that all proximate determinants are very strongly related to infants' survival. In particular, a short interval between succeeding births is correlated with a substantially increased likelihood of infant death. Children born to women younger than 20 years have worse survival chances than those born to women between 20 and 35 years. Firstborn children and children of higher birth order (above four) have a higher probability of dying within their first year.

Regarding the socioeconomic determinants, we find that maternal education reduces the risk of infant mortality. The point estimate is larger in rural areas but the difference is not significant. All proxies for traditions, social norms and attitudes, except for mother's age at first marriage, are highly correlated with infant mortality. Familiarity with contraception, an older household head and, in rural areas, a female head of the household are all positively correlated with infants' survival. Female children have better survival chances than their male counterparts.[17]

Household environmental conditions, in particular access to potable water, appear to be important determinants of infant mortality risks. In rural locations, the very few households with an electricity supply have a greatly reduced probability of infant death. In urban areas, the mortality risk is substantially higher among households living in premises with no finished floor. It seems likely that this characteristic identifies slum dwellings and the poor public health conditions found there. In rural areas, the majority of dwellings have no finished floor and this is not significantly correlated with mortality risk. Surprisingly, having a toilet is not significantly correlated with mortality risk in either urban or rural areas. Children in households with fewer assets face a greater risk of death in urban but not in rural areas. This is consistent with a greater socioeconomic gradient in child health in urban areas that has been found in other studies (Fotso 2006; Van de Poel *et al* 2007).

At the community level, the existence of a health facility is correlated with a reduced risk of death but the availability of public transport has no significant effect. Infant mortality is highest in Mali and lowest in CAR, although in rural areas there are no differences between Benin, Chad and CAR. Unobserved household level heterogeneity explains 20% of the variance in infant mortality that remains after controlling for observable determinants, whereas community level heterogeneity is significant but small (0.8%). The relative importance of the household level variance is evident from the strong clustering of deaths by household seen in Table 2. Curtis *et al* (1993) also found household heterogeneity explaining about 23% of the random variance in infant mortality in Brazil. However, another study of child survival (to age five rather than one) in Brazil that allowed for both household and community random effects found the latter to be more important (Sastry 1997b). The only other study of child survival that has allowed for both effects was of Malawi in sub-Saharan Africa and this, like the present study, found household level heterogeneity to be more important (Bolstad & Manda 2001). Household and community level heterogeneity are largest within rural areas. The community component is even absent within urban areas.

When using under-five instead of infant mortality, which almost doubled the number of deaths, we still found a very small estimate of community level variance. This suggests that the low community level variance is not just due to the smaller number of deaths in urban communities.[18] Further, the community level variance did not increase much by omitting the household random effect, suggesting that there is not a problem of separately identifying the two effects. Finally, when we re-estimated the model omitting community level covariates, the community level variance did not increase by much suggesting that it is not the case that there is a large community level effect that is adequately captured by observable characteristics.

Decomposition results

The decomposition method (6) reveals that 100% of the rural-urban gap in infant mortality can be explained by differences in the distributions of the covariates and the random effects and so, in aggregate, differences in the coefficients do not explain any of the gap. This does not mean that there are no differences in the effects of determinants of infant mortality across rural and urban areas. Rather, there are no systematic differences.

Some determinants have a stronger effect in rural areas, for example electricity supply, while others have a stronger effect in urban areas, for example possession of assets. Given the limited evidence we found of significant rural-urban differences in coefficients and their zero net effect in aggregate, in the remainder of the analysis we focus on the contributions of differences in the distributions of observable and unobservable determinants of infant mortality.

The absolute and relative contributions of each covariate and of the random effects are given in the first two columns of Table 7 and the relative contributions are presented in Figure 2.[19] The contribution of a variable derives from the difference between the rural and urban distributions of the variable and from the magnitude of its association with infant mortality (in the pooled model) as given in the first column of Table 6. Proximate determinants contribute only about 7% of the rural-urban gap, whereas socioeconomic determinants account for 56% of the gap. The contribution of proximate determinants derives mainly from rural-urban differences in the prevalence of short birth intervals (6%). Within the socioeconomic characteristics, the most important contribution comes from environmental conditions, with water supply, electricity supply and quality of housing, as indicated by flooring materials, each accounting for 12-13% of the gap. Maternal education accounts for about 6% of the gap. Except for familiarity with contraception, which contributes 5%, all other proxies for traditions, social norms and attitudes do not contribute much. Differences in household level unobserved heterogeneity contributes a substantial 10% to the gap.

Community characteristics contribute 13% to the gap, the most important contribution coming from the existence of a health facility (9%). Unobserved community level heterogeneity is only responsible for 2% of the gap. The contribution of the country effects amounts to 14% and is caused by 2 factors. First, there are differences across countries in the urban/rural population split (Table 1), and therefore the proportion of infants from any one country in the pooled sample differs across urban and rural areas. Second, infant mortality differs across countries even after controlling for all covariates (Table 6).

The main difference between these decomposition results and others derived without taking account of unobservable heterogeneity is in the contribution of potable water, which is about 9% points higher in the latter decomposition (results are available upon request). This suggests that water supply acts as a proxy for unobservable determinants when there is no explicit allowance for household and community level heterogeneity.

Country specific analysis

The effects of covariates and the size of the household and community effects may differ across countries. We therefore carry out the analysis for each country separately. Table 8 shows regression results for each country. Proximate determinants in the form of maternal characteristics have very consistent effects on infant mortality. In particular, the effect of a short birth interval is consistently significant and large. This consistency presumably reflects both the biological nature of the relationships and the limited scope for cross-country variability in measurement of the variables. The effects of the socioeconomic characteristics display much more cross country variation. Kuate-Defo & Diallo (2002) found that the effects of bio-demographic variables are also much more stable across countries and time than those of socioeconomic covariates. Notwithstanding this variation, the country-specific results generally confirm the results from the pooled analysis. The lower significance of coefficients in the country specific models is partly due to much lower sample sizes compared to the pooled analysis.

Among the household socioeconomic variables, assets generally have no significant effect on infant mortality. Mother's education is associated with a reduced risk of death but the effect is significant only in Mali, Guinea and rural Chad. Other studies have also found a small effect of maternal education on child health in Sub-Saharan Africa (Hobcraft 1993; Adetunji 1995; Lalou & LeGrand 1997; Hong 2006). The household environmental variables generally take the expected signs and water supply has the most consistent effect, significantly reducing the risk of death in four of the six countries. Proximity to health services is usually associated with a reduced risk of infant mortality but the effect reaches significance only in Niger, Guinea and urban CAR. The public transport effect is generally negative but not significant.

As in the pooled model, unobserved household heterogeneity explains a substantial part of the random variance in infant mortality: 8% in Mali, 13% in Guinea, 24% in Benin, 28% in Niger and 33% in Chad. The results for CAR, where the household level variance is zero, are somewhat out of line with the others. This is partly due to the small sample size and the relatively low number of deaths, which makes it difficult to identify the unobserved components. Repeating the analysis for under-five rather than infant mortality did yield a larger household level variance in this case. The proportion of the random variance explained by community level heterogeneity is very small. It is highest in Niger at 2.6% and insignificant in Benin and CAR. As in the pooled model, community level heterogeneity is much larger within rural areas (except in Benin). These results were also found to be robust to switching to under-five mortality, omitting the household level heterogeneity and using a linear probability model.

Again, coefficients generally do not differ much across urban/rural samples. However, in Niger and CAR the health benefits of clean water are larger in urban areas and in Chad they are larger in rural areas. Electricity has a stronger positive effect on health in the rural areas of Benin and Guinea. There are also some cases of rural-urban differences in the effects of the proximate determinants.

The decomposition (6) shows that the major part of the rural-urban gap is caused by differences in the distributions of determinants (Mali: 102%, Benin: 87%, CAR: 103%, Chad: 113%, Niger: 102%, Guinea: 95%). Table 7 gives detailed decomposition of this part of the gap for each country. In general, these country specific results are in line with the pooled analysis, with the major contributions coming from household environmental characteristics (ranging from 22% in Chad to 57% in Guinea). With the exception of contraceptive use, other proxies for social norms and attitudes are generally not so important in accounting for the gap. Mum's education accounts for 4% of the gap in CAR and up to 9% in Guinea. The contributions of the proximate determinants are generally much smaller than those of the socioeconomic determinants. Their total contribution ranges from 4.5% in Chad to 25% in Benin and is mainly driven by rural-urban differences in short birth intervals. The contribution of unobserved household heterogeneity displays substantial cross-country variation, from less than 2% in Mali to 60% in Chad. The contribution of community level characteristics seems to be mainly driven by availability of health care, although in Mali public transport is also playing a role. The contribution of community level heterogeneity is fairly small, ranging from 0.24% in CAR to 5% in Niger.

CONCLUSION

The rural-urban gap in infant mortality in six West sub-Saharan African countries is explained by differences in the distributions of factors that determine mortality and not by differences in the effects of those determinants between rural and urban locations. Almost three-quarters of the gap is explained by differences in the distributions of household level factors. Most of this household level contribution is accounted for by observable covariates but 10% of the gap is due to unobservable household level determinants. Rural-urban differences in the distributions of community level determinants are much less important. In total, they explain 13% of the gap in infant mortality, including 3% that derives from unobservable community level heterogeneity. Among the observable household level determinants, environmental factors make the greater contribution to explanation of the rural-urban gap in infant mortality rates. Access to potable water, to electricity and the quality of housing materials together explain 37% of the gap. Conditional on these environmental factors, differences in socio-economic status explain around 8% of the gap with most of this due to the effect of mothers' education while the possession of assets explains little more than 1% of the gap.

Although maternal characteristics are very strong and consistent proximal determinants of infant mortality, they contribute relatively little to the rural-urban gap since their distributions differ little between rural and rural areas. One exception is the greater prevalence of short birth intervals in rural areas that accounts for almost 7% of the gap. Of the factors intended to pick up the effects of differences in cultures and norms, only the lower use of contraception in rural areas makes a substantial contribution, explaining 6% of the gap.

Most of the contribution of community level factors is due to the lower proximity to health facilities in rural areas, which explains 9% of the gap. Unobserved community level heterogeneity is contributing only marginally to the random variance in infant mortality, particularly in urban areas, and to the rural-urban gap. These results are robust to replacing infant with under-five mortality and omitting the household level heterogeneity.

Allowing for unobserved heterogeneity in the decomposition is important not only because it reveals the contribution made by differences in household and community level unobservable determinants to the rural-urban gap in infant mortality, but also because accounting for them provides better estimates of the contribution of the observed characteristics. This is illustrated by the substantial reduction in the contribution of safe drinking water after accounting for heterogeneity, which suggests that it is acting as a proxy for unmeasured household/community mortality risk in models that do not take account of such heterogeneity. As would be expected, the contributions of variables that are essentially independent among siblings, such as child sex, are the most robust to inclusion of the random intercepts.

In sum, this paper shows that child survival depends first and foremost on the living conditions that constrain the ability of households to care for their children. It is not so much that rural households behave differently from their urban counterparts, but that they live under conditions that are more detrimental to their infants' health. The decomposition reveals that the larger part of the rural-urban gap in infant mortality is caused by differences at the household rather than the community level. This suggests that policy interventions would be most effective in reducing the excess rural infant mortality if targeted at disadvantaged households instead of at complete areas. Of course, while environmental factors, such as electricity and water supply, do vary across households within communities, access to such services is clearly not independent from the community level infrastructure (Sastry 1996). The large contribution of these factors to the rural-urban gap in infant mortality suggests that community level interventions are certainly necessary to reduce the gap.

REFERENCES

Adetunji, J.A. 1995. "Infant mortality and mothers education in Ondo state, Nigeria." *Social Science and Medicine* 40(2): 253-263.

Bhargava, A. 2003. "Family planning, gender differences and infant mortality: evidence from Uttar Pradesh, India." *Journal of Econometrics* 112: 225-240.

Birdsall, N. & L.A. Chester. 1987. "Contraception and the status of women: what is the link?" *Family Planning Perspectives* 19 (1): 14-18.

Blinder, A.S. 1973. "Wage Discrimination: Reduced Form and Structural Estimates." *Journal of Human Resources* 8: 436-455.

Brenneman, A. 2002. *Infrastructure & poverty linkages: A literature review.* Washington, DC: The World Bank.

Brockerhoff, M. 1995. "Child survival in big cities: the disadvantages of migrants." *Social Science & Medicine* 40: 1371-1383

Bolstad, W.M. & S.O. Manda. 2001. "Investigating child mortality in Malawi using family and community random effects." *Journal of the American Statistical Association* 96 (453): 12-19.

Cai, L. & V. Chongsuvivatwong. 2006. "Rural-urban differentials of premature mortality burden south-west China." *International Journal for Equity in health* 5.

Caldwell, J. 1979. "Education as a factor in mortality decline: An examination of Nigerian data." *Population Studies* 33 (3): 395-413.

Canagarajah, N. 2001. "Child Labor in Africa: A Comparative Study." *The ANNALS of the American Academy of Political and Social Science* 575: 71-91

Cleland, J.G. & J.K. van Ginneken. 1988. "Maternal education and child survival in developing countries: the search for pathways of influence." *Social Science & Medicine* 27 (12): 1357–1368.

Cleland, J., J. Bicego & G. Fegan. 1992. "Socioeconomic inequalities in childhood mortality: the 1970s to the 1980s." *Health Transition Review* 2: 1-18.

Curtis, S.L., I. Diamond & J.W. McDonald. 1993. "Birth interval and family effects on postneonatal mortality in Brazil." *Demography* 30(1): 33-43.

Esrey, S.A., J.B. Potash, L. Roberts & C. Shiff. 1991. "Effects of improved water supply & sanitation on ascariasis, diarrhea, dracunculiasis, hookworm infection, schistosomiasis, & trachoma." *Bulletin of the World Health Organization* 89(5):609–621.

Fairlie, R.W. 2005. "An extension of the Blinder-Oaxaca decomposition technique to logit and probit models." *Journal of Economic and Social Measurement* 30: 305-316.

Filmer, D., & L. Pritchett. 2001. "Estimating wealth effects without expenditure data - or tears: An application to educational enrolments in states of India." *Demography* 38: 115-132.

Folasada, I.B. 2000. "Environmental factors, situation of women and child mortality in southwestern Nigeria." *Social Science and Medicine* 51(10): 1473-1489.

Fotso, J-C. 2006. "Child health inequities in developing countries: differences across urban and rural areas." *International Journal for Equity in Health* 5: 9.

Gibbons, R. & D. Hedeker. 1997. "Random effects probit and logistic regression models for three level data." *Biometrics* 53: 1527-37.

Gould, W. 1998. “African mortality and the new ‘urban penalty’.” *Health and Place* 4(2): 171-181.

Heaton, T.B., & R. Forste. 2003. “Rural-urban differences in child growth and survival in Bolivia.” *Rural Sociology* 68 (3): 410-433.

Hertz, E., J.R. Herbert & J. Landon. 1994. “Social and environment factors and life expectancy, infant mortality, and maternal mortality rates: results of a cross-national comparison.” *Social Science & Medicine* 39 (1): 105–114.

Hobcraft, J.M. 1993. “Women’s education, child welfare and child survival: a review of evidence.” *Health Transition Review* 3 (3): 159–175.

Hong, R. 2006. “Effect of economic inequality on chronic childhood undernutrition in Ghana.” *Public Health Nutrition* 10: 371-378.

Kalipeni, E. 1993. “Determinants of infant mortality in Malawi: a spatial perspective.” *Social Science & Medicine* 39 (2): 183–198.

Knobel, H.H., W.S. Yang & M.S. Ho. 1994. “Urban-rural and regional differences in infant mortality in Taiwan.” *Social Science & Medicine* 39 (6): 815–822.

Kuate-Defo, B. & K. Diallo, K. 2002. “Geography of child mortality clustering within African families.” *Health & Place* 8: 93-117.

Lalou, R. & T. LeGrand. 1997. “Child mortality in the Urban and Rural Sahel.” *Population: an English Selection* 9: 147-168.

Lloyd, C.B. & A.K. Blanc. 1996. “Children's schooling in sub-Saharan Africa: the role of fathers, mothers, and others.” *Population and Development Review* 22 (2): 265–298.

Manda, S. 1999. “Birth intervals, breastfeeding and determinants of childhood mortality in Malawi.” *Social Science and Medicine* 48: 301-312.

Menon, P., M. Ruel & S. Morris. 2000. “Socio-economic differentials in child stunting: Results from 11 DHS data sets.” *Food and Nutrition Bulletin* 21(3): 282-289.

Molbak, K., H. Jensen, L. Ingholt & P. Aaby. 1997. “Risk factors for diarrhoeal disease incidence in early childhood.” *American Journal of Epidemiology* 146 (3): 273-282.

Mosley, W.H. & L.C. Chen. 1984. “An analytical framework for the study of child survival in developing countries.” *Population and Development Review* 10: 25-45.

Neumark, D. 1988. “Employers' discriminatory behavior and the estimation of wage discrimination.” *Journal of Human Resources* 23: 279-295.

Oaxaca, R. 1973. “Male-female wage differentials in urban labor markets.” *International Economic Review* 14: 693-709.

Oaxaca, R., & M. Ransom. 1994. “On Discrimination and the Decomposition of Wage Differentials.” *Journal of Econometrics* 61: 5-21.

Rabe-Hesketh, S, A. Skrondal & A. Pickles. 2005. “Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. “ *Journal of Econometrics* 128: 301-323.

Rabe-Hesketh, S., A. Skrondal & A. Pickles. 2002. “Reliable estimation of generalized linear mixed models using adaptive quadrature.” *The Stata Journal* 2 (1): 1-21.

Ronsmans, C. 1996. “Birth spacing and child mortality in rural Senegal.” *International Journal of Epidemiology* 25: 989-997.

Rutstein, S.O. 2000. "Factors associated with trends in infant & child mortality in developing countries during the 1990s." *Bulletin of the World Health Organization* 78(10): 1256–1270.

Sastry, N. 1996. "Community characteristics, individual and household attributes, and child survival in Brazil." *Demography* 33 (2): 211-229.

_________1997a. "What explains rural-urban differentials in child mortality in Brazil?" *Social Science and Medicine* 44(7): 989-237.

_________1997b. "A Nested Frailty model for survival data, with an application to the study of child survival in northeast Brazil." *Journal of the American Statistical Association* 92 (438): 426-435.

Rosenzweig, M.R. & P.T. Schultz. 1982. "Market opportunities, genetic endowments, and intrafamily resource distribution: child survival in rural India." *The American Economic Review* 72 (4): 803-815.

Smith, L., M. Ruel & A. Ndiaye. 2005. "Why is child malnutrition lower in urban than in rural areas? Evidence from 36 developing countries." *World Development* 33(8): 1285-1305.

Tambiah, S.J., M. Goheen, A. Gottlieb, J.I. Guyer, E.A. Olson, C. Piot, K.W. Van Der Veen & T. Vuyk. 1989. "Bridewealth and dowry revisited: The position of women in Sub-Saharan Africa and North India." *Current Anthropology* 20 (4): 413-435.

Tantchou, J & E. Wilson. 2000. *Post-Cairo reproductive health policies and programs: A study of five Francophone African Countries.* Washington: POLICY Project.

Van de Poel, E., O. O'Donnell & E. van Doorslaer. 2007. "Are urban children really healthier? Evidence from 47 developing countries." *Social Science and Medicine*, forthcoming.

Victora, C., B. Fenn, J. Bryce & B. Kirkwood. 2005. "Co-coverage of preventive interventions and implications for child-survival strategies: evidence from national surveys." *The Lancet* 366: 1460-1466.

Wang, L. 2003. "Determinants of child mortality in LDCs. Empirical findings from demographic and health surveys." *Health Policy* 65: 277-299.

Wooldridge, J.M. 2002. *Econometric analysis of Cross Section and Panel Data*. Cambridge: MIT Press, 752 p.

World Bank. 2006. World Development Indicators 2006. Washington, DC: The World Bank.

World Health Organization. 2002. The World Health Report 2002: Reducing risks, promoting healthy life. Geneva: The World Health Organization.

country	year of survey	sample size	infant mortality rate	urban infant mortality rate	rural infant mortality rate	rural-urban gap	urban population (% of total)
Benin	2001	8124	0.1023	0.0780	0.1122	**0.0341**	28.87
CAR	1995	5436	0.0964	0.0781	0.1075	**0.0294**	38.03
Chad	2004	8738	0.1203	0.1037	0.1239	**0.0202**	18.06
Gabon	2000	6234	0.0631	0.0624	0.0648	0.0023	71.17
Guinea	1999	9403	0.1276	0.0901	0.1393	**0.0491**	23.67
Mali	2001	20176	0.1515	0.1232	0.1588	**0.0355**	20.32
Niger	1998	10596	0.1503	0.0936	0.1604	**0.0668**	15.15
pooled		62473	0.1315	0.0959	0.1414	**0.0454**	21.67

Table 1: Rural and urban infant mortality rates estimated from Demographic and Health Surveys.
NOTE: Infant mortality rates are the proportion of infant deaths per 100 live births.
Rural-urban gaps in bold indicate significance at the 10% level.
The pooled estimates are weighted to reflect the different population sizes of countries.

		URBAN														RURAL															
		deaths per household														deaths per household															
children per household	(%)	0		1		2		3		4		5		Total	(%)	0		1		2		3		4		5		>6		Total	(%)
1	(%)	1062	(93)	79	(7)									1141	(20)	2002	(88)	278	(12)											2280	(16)
2	(%)	1238	(88)	158	(11)	9	(1)							1405	(25)	2563	(82)	512	(16)	62	(2)									3137	(22)
3	(%)	1243	(79)	279	(17)	46	(3)	5	(0)					1573	(28)	2914	(72)	948	(23)	163	(4)	12	(0)							4037	(28)
4	(%)	613	(67)	236	(26)	61	(7)	3	(0)	1	(0)			914	(16)	1477	(55)	888	(33.)	251	(9)	57	(2)	7	(0)					2680	(18)
5	(%)	184	(55)	91	(27)	41	(12)	13	(4)	5	(1)	0	(0)	334	(6)	466	(43)	352	(32)	199	(18)	58	(5)	13	(1)	1	(0)			1089	(8)
6	(%)	76	(53)	37	(26)	18	(13)	9	(6)	2	(1)	1	(1)	143	(3)	201	(38)	179	(34)	87	(17)	39	(7)	13	(2)	4	(1)	0	(0)	523	(4)
7	(%)	41	(48)	30	(35)	9	(10)	5	(6)	1	(1)	0	(0)	86	(1)	101	(32)	113	(35)	61	(19)	27	(8)	8	(3)	7	(2)	1	(0)	318	(2)
8	(%)	19	(38)	15	(30)	11	(22)	4	(8)	1	(2)	0	(0)	50	(1)	66	(30)	64	(29)	46	(21)	23	(11)	11	(5)	5	(2)	2	(1)	217	(1)
9	(%)	12	(44)	8	(30)	2	(7)	2	(7)	2	(7)	1	(4)	27	(0)	22	(22)	28	(29)	17	(17)	17	(17)	10	(10)	3	(3)	1	(1)	98	(1)
10	(%)	1	(6)	5	(31)	7	(44)	0	(0)	0	(0)	3	(19)	16	(0)	12	(22)	11	(20)	12	(22)	9	(17)	7	(13)	2	(4)	1	(2)	54	(0)
11	(%)	3	(60)	0	(0)	0	(0)	1	(20)	1	(20)	0	(0)	5	(0)	3	(14)	5	(23)	6	(27)	5	(23)	1	(5)	1	(5)	1	(5)	22	(0)
12	(%)	0	(0)	3	(100)	0	(0)	0	(0)	0	(0)	0	(0)	3	(0)	3	(18)	1	(6)	6	(35)	1	(6)	5	(29)	1	(6)	0	(0)	17	(0)
13	(%)	0	(0)	0	(0)	2	(100)	0	(0)	0	(0)	0	(0)	2	(0)	0	(0)	1	(14)	2	(29)	2	(29)	2	(29)	0	(0)	0	(0)	7	(0)
>14	(%)									1	(100)			1	(0)	3	(25)	0	(0)	1	(8)	0	(0)	4	(33)	2	(17)	2	(17)	12	(0)
Total hhold	(%)	4492	(79)	941	(17)	206	(4)	42	(1)	14	(0)	5	(0)	5700	(100)	9833	(68)	3380	(23)	913	(6)	250	(2)	81	(1)	26	(0)	8	(0)	14491	(100)
Total deaths	(%)		(0)	941	(60)	412	(26)	126	(8)	56	(4)	25	(2)	1560	(100)	0	(0)	3380	(52)	1826	(28)	750	(12)	324	(5)	130	(2)	56	(1)	6466	(100)

Table 2: Distribution of children and infant deaths across households.

URBAN																													
		deaths per community																											
children per comm	(%)	0		1		2		3		4		5		6		7		8		9		10		10-15		>15		Total	(%)
<20	(%)	110	(44)	73	(29)	41	(17)	14	(6)	10	()	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)			248	(38)
20-30	(%)	30	(16)	52	(28)	39	(21)	31	(17)	15	(8)	8	(4)	6	(3)	2	(1)	2	(1)	1	(1)	0	(0)	0	(0)			186	(29)
30-40	(%)	12	(11)	16	(14)	22	(19)	13	(11)	17	(15)	14	(12)	6	(5)	10	(9)	3	(3)	1	(1)	0	(0)	0	(0)			114	(17)
40-50	(%)	3	(4)	4	(6)	5	(7)	14	(20)	10	(14)	10	(14.)	9	(13)	6	(9)	4	(6)	2	(3)	2	(3)	1	(1)			70	(11)
50-60	(%)	0	(0)	0	(0)	1	(4)	3	(12)	3	(12)	3	(12)	4	(16)	2	(8)	2	(8)	4	(16)	0	(0)	3	(12)			25	(4)
60-70	(%)	0	(0)	0	(0)	0	(0)	0	(0)	1	(33)	0	(0)	0	(0)	2	(67)	0	(0)	0	(0)	0	(0)	0	(0)			3	(0)
70-80	(%)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	1	(50)	0	(0)	1	(50)	0	(0)			2	(0)
90-100	(%)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	1	(100)			1	(0)
>100	(%)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	1	(100)			1	(0)
Total comm	(%)	155	(24)	145	(22)	108	(17)	75	(11)	56	(9)	35	(5)	25	(4)	22	(3)	12	(2)	8	(1)	3	(0)	6	(1)			650	(100)
Total deaths	(%)	0	(0)	145	(9)	216	(14)	225	(14)	224	(14)	175	(11)	150	(10)	154	(10)	96	(6)	72	(5)	30	(2)	73	(5)			1560	(100)

RURAL																													
		deaths per community																											
children per comm	(%)	0		1		2		3		4		5		6		7		8		9		10		10-15		>15		Total	(%)
<20	(%)	23	(26)	22	(25)	25	(28)	10	(11)	6	(7)	1	(1)	1	(1)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	88	(9)
20-30	(%)	9	(7)	26	(19)	29	(21)	22	(16)	23	(17)	15	(11)	7	(5)	3	(2)	0	(0)	1	(1)	0	(0)	0	(0)	0	(0)	135	(14)
30-40	(%)	8	(5)	9	(6)	19	(12)	20	(13)	23	(15)	19	(12)	14	(9)	12	(8)	15	(10)	10	(6)	3	(2)	4	(23)	1	(1)	157	(16)
40-50	(%)	1	(1)	4	(2)	18	(11)	16	(9)	18	(11)	27	(16)	22	(13)	19	(11)	11	(6)	13	(8)	7	(4)	10	(6)	4	(2)	170	(18)
50-60	(%)	0	(0)	2	(1)	3	(2)	7	(5)	10	(7)	12	(9)	14	(10)	10	(7)	20	(14)	22	(16)	15	(11)	21	(15)	3	(2)	139	(14)
60-70	(%)	1	(1)	0	(0)	2	(2)	7	(6)	7	(6)	9	(8)	10	(9)	14	(13)	12	(11)	6	(6)	6	(6)	26	(24)	9	(8)	109	(11)
70-80	(%)	0	(0)	0	(0)	3	(4)	5	(6)	0	(0)	7	(8)	3	(4)	10	(12)	4	(5)	6	(7)	5	(6)	23	(27)	19	(22)	85	(9)
80-90	(%)	0	(0)	0	(0)	2	(5)	0	(0)	0	(0)	1	(2)	2	(5)	1	(2)	1	(2)	1	(2)	5	(12)	17	(40)	13	(30)	43	(4)
90-100	(%)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	2	(9)	1	(5)	6	(27)	13	(60)	22	(2)
>100	(%)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)	1	(6)	1	(6)	2	(12)	13	(76)	17	(2)
Total comm	(%)	42	(4)	63	(7)	101	(10)	87	(9)	87	(9)	91	(9)	73	(8)	69	(7)	63	(7)	62	(6)	43	(4)	109	(11)	75	(8)	965	(100)
Total deaths	(%)	0	(0)	63	(1)	202	(3)	261	(4)	348	(5)	455	(7)	438	(7)	483	(7)	504	(8)	558	(9)	430	(7)	1322	(20)	1402	(22)	6466	(100)

Table 3: Distribution of children and infant deaths across communities.

		Variable	Description
Proximate determinants	***maternal factors***	firstborn	1 if child is mum's firstborn, 0 otherwise
		birth order>4	1 if child's birth order is higher than four, 0 otherwise
		mum's age at birth	3 categories: <20, [20,35], >35
		short birth interval	1 if less than 12 months between preceding birth, 0 otherwise
Socioeconomic determinants	***education***	mum not educated	1 if mum had no education, 0 otherwise
	traditions / social norms / attitudes	contraception	1 if mum has ever used modern contraception, 0 otherwise
		age 1st marriage	mum's age at her first marriage (in years)
		sex of child	1 if child is male, 0 otherwise
		age of household head	in years
		sex of household head	1 if head is male, 0 otherwise
	environmental	toilet	1 if household disposes of any toilet facility, 0 otherwise
		water	1 if water coming from tap, protected well, bottle, vendor, 0 otherwise
		electricity	1 if household has electricity access, 0 otherwise
		no finished floor	1 if household has no finished floor (sand or mud), 0 otherwise
	economic status	assets	3 categories: poorest (asset1), middle (asset2), asset3 (richest)
Community determinants	***service and infrastructure characteristics***	health facilities	1 if health facility is within community, 0 otherwise
		public transport	1 if there is some form of public transport within the community, 0 otherwise

Table 4: Description of covariates. Underscored variables are the reference category used in regressions.

		POOLED		MALI		BENIN		CAR		CHAD		NIGER		GUINEA	
		urban	rural	urban	rural	urban	rural	urban	rural	urban	rural	urban	rural	urban	rural
Proximate determinants	firstborn	**0.21**	**0.18**	**0.20**	**0.16**	**0.26**	**0.19**	0.22	0.21	**0.21**	**0.19**	0.19	0.17	**0.22**	**0.18**
	birth order>4	**0.32**	**0.38**	**0.35**	**0.41**	**0.28**	**0.35**	0.30	0.29	0.35	0.36	**0.38**	**0.41**	**0.28**	**0.35**
	mum's age at birth<20	**0.22**	**0.22**	0.22	0.22	**0.15**	**0.18**	**0.25**	**0.21**	0.25	0.25	**0.21**	**0.25**	0.22	0.22
	35<mum's age at birth>20	**0.70**	**0.67**	**0.69**	**0.66**	**0.74**	**0.71**	**0.69**	**0.71**	**0.68**	**0.66**	**0.73**	**0.67**	0.69	0.68
	mum's age at birth>35	**0.08**	**0.10**	**0.09**	**0.12**	0.11	0.11	**0.06**	**0.08**	**0.07**	**0.08**	**0.06**	**0.08**	**0.09**	**0.10**
	short birth interval	**0.21**	**0.25**	**0.23**	**0.28**	**0.14**	**0.19**	**0.22**	**0.28**	0.27	0.27	**0.23**	**0.25**	**0.16**	**0.19**
Socioeconomic determinants	mum not educated	**0.81**	**0.97**	**0.84**	**0.98**	**0.82**	**0.97**	**0.74**	**0.95**	**0.81**	**0.99**	**0.77**	**0.96**	**0.80**	**0.97**
	contraception	**0.31**	**0.11**	**0.41**	**0.18**	**0.30**	**0.18**	**0.26**	**0.06**	**0.14**	**0.03**	**0.43**	**0.08**	**0.22**	**0.07**
	age 1st marriage	**16.78**	**15.98**	**16.71**	**16.14**	**18.71**	**17.50**	**16.33**	**16.90**	15.76	15.70	**15.73**	**14.68**	**16.90**	**15.98**
	sex of child	0.51	0.51	**0.52**	**0.50**	0.50	0.50	0.50	0.51	0.52	0.51	0.51	0.51	0.52	0.51
	age of household head	42.74	42.61	**42.32**	**42.85**	**41.59**	**42.75**	**40.40**	**38.32**	**42.78**	**39.59**	**43.39**	**42.77**	**46.39**	**46.96**
	sex of household head	**0.87**	**0.92**	**0.91**	**0.93**	**0.85**	**0.91**	**0.83**	**0.92**	**0.85**	**0.90**	0.94	0.93	**0.85**	**0.93**
	toilet	**0.85**	**0.40**	**0.98**	**0.72**	**0.52**	**0.12**	**0.95**	**0.61**	**0.81**	**0.14**	**0.78**	**0.06**	**0.94**	**0.50**
	water	**0.71**	**0.26**	**0.66**	**0.34**	**0.65**	**0.33**	**0.85**	**0.45**	**0.70**	**0.31**	**0.92**	**0.12**	**0.57**	**0.05**
	electricity	**0.31**	**0.02**	**0.32**	**0.02**	**0.41**	**0.04**	**0.09**	**0.01**	**0.18**	**0.00**	**0.33**	**0.00**	**0.48**	**0.02**
	no finished floor	**0.41**	**0.88**	**0.42**	**0.94**	**0.25**	**0.56**	**0.67**	**0.94**	**0.79**	**0.99**	**0.32**	**0.97**	**0.10**	**0.75**
	asset1	**0.16**	**0.34**	**0.13**	**0.29**	**0.19**	**0.33**	**0.18**	**0.30**	**0.12**	**0.43**	**0.17**	**0.35**	**0.19**	**0.36**
	asset2	**0.26**	**0.36**	**0.23**	**0.36**	**0.32**	**0.36**	0.31	0.33	**0.28**	**0.38**	**0.26**	**0.37**	**0.24**	**0.36**
	asset3	**0.58**	**0.30**	**0.64**	**0.35**	**0.49**	**0.31**	**0.51**	**0.37**	**0.61**	**0.19**	**0.58**	**0.28**	**0.57**	**0.28**
Community determinants	health facilities	**0.71**	**0.23**	**0.91**	**0.28**	**0.57**	**0.35**	**0.57**	**0.18**	**0.77**	**0.11**	**0.74**	**0.12**	**0.53**	**0.28**
	public transport	**0.55**	**0.27**	**0.70**	**0.20**	**0.36**	**0.16**	**0.60**	**0.50**	**0.53**	**0.09**	**0.97**	**0.58**	**0.12**	**0.23**

Table 5: Urban/rural means of covariates.
NOTE: Bold numbers indicate that urban and rural mean differ significantly at the 10% level.

		POOLED		
	Variables	total	urban	rural
Proximate determinants	firstborn	0.355***	0.261***	0.381***
	birth order>4	0.138***	0.249***	0.115**
	mum's age at birth<20	0.371***	0.409***	0.359***
	mum's age at birth>35	-0.080	0.050	-0.104
	short birth interval	0.551***	0.658***	0.531***
Socioeconomic determinants	mum not educated	0.277***	0.183*	0.321**
	contraception	-0.233***	-0.314***	-0.189***
	age 1st marriage	-0.003	-0.002	-0.002
	sex of child	0.148***	0.112*	0.155***
	age of household head	0.003**	0.003	0.003**
	sex of household head	**0.159*****	**-0.036**	**0.240*****
	toilet	**-0.033**	**0.141**	**-0.061**
	water	-0.165***	-0.183**	-0.153***
	electricity	**-0.269*****	**-0.098**	**-0.405****
	no finished floor	**0.151*****	**0.255*****	**0.060**
	asset1	**0.054**	**0.277*****	**0.017**
	asset2	**0.014**	**0.140***	**-0.018**
Community determinants	health facilities	-0.117***	-0.195***	-0.115**
	public transport	-0.040	0.044	-0.084
Country fixed effects	Benin	0.125	0.228	0.066
	Chad	**0.055**	**0.272****	**-0.041**
	Guinea	0.292***	0.326**	0.265***
	Mali	0.507***	0.654***	0.466***
	Niger	0.351***	0.400***	0.338***
	constant	-3.427***	-3.425***	-3.405***
	observations	62473	16396	46077
	variance of hh effect	0.771***	0.384***	0.850***
	variance of comm effect	0.073***	0.000	0.083***
	mean prediction	0.132	0.096	0.141

Table 6: Regression coefficients for the total, urban and rural sample. Based upon data pooled across all countries.

NOTE: Significance at: * 10%, ** 5%, * 1%.**

Coefficients in bold indicate that they differ significantly between urban and rural model at the 10% level.

		POOLED		MALI		BENIN		CAR		CHAD		NIGER		GUINEA	
	Variables	contri-bution	%	con	%	con	%	con	%	con	%	con	%	con	%
Proximate determinants	firstborn	-0.09	-2.03	-0.46	-11.90	-0.06	-2.25	-0.63	-21.13	-0.08	-2.01	0.08	1.17	-0.07	-1.61
	birth order>4	0.05	1.17	0.17	4.45	0.21	7.96	0.10	3.35	-0.02	-0.62	0.02	0.24	0.17	3.81
	mum's age at birth	0.09	1.99	0.12	3.04	0.16	5.90	0.00	-0.06	0.00	-0.05	0.30	4.23	0.06	1.32
	short birth interval	0.29	6.25	0.78	20.34	0.36	13.62	0.89	29.61	0.27	7.22	0.06	0.84	0.14	3.15
Socioeconomic determinants	mum not educated	0.30	6.37	0.33	8.59	0.19	7.36	0.11	3.72	0.25	6.48	0.33	4.62	0.41	9.21
	contraception	0.26	5.46	0.38	9.76	0.01	0.39	-0.09	-2.84	0.07	1.74	1.08	15.28	0.33	7.41
	age 1st marriage	0.01	0.30	0.11	2.82	-0.05	-2.00	-0.02	-0.78	0.00	-0.09	0.07	1.06	-0.16	-3.47
	sex of child	0.02	0.40	0.04	0.96	0.00	-0.07	0.04	1.46	0.02	0.53	0.02	0.28	0.02	0.43
	age of household head	-0.02	-0.45	-0.01	-0.21	0.02	0.72	-0.16	-5.35	0.03	0.71	-0.06	-0.84	0.00	0.09
	sex of household head	0.08	1.67	0.12	3.05	-0.01	-0.36	0.34	11.31	0.03	0.70	0.06	0.90	-0.09	-1.94
	toilet	0.11	2.36	0.16	4.24	0.51	19.38	0.05	1.70	-1.10	-29.06	-0.82	-11.57	0.38	8.35
	water	0.61	13.08	0.07	1.82	0.17	6.63	0.84	27.89	1.40	36.97	1.22	17.36	0.97	21.57
	electricity	0.58	12.45	0.34	8.69	0.55	20.76	0.07	2.29	0.01	0.36	1.67	23.65	0.59	13.20
	no finished floor	0.59	12.51	1.11	28.93	0.19	7.36	0.67	22.23	0.52	13.70	0.37	5.28	0.61	13.64
	assets	0.07	1.45	0.18	4.61	-0.01	-0.26	0.21	6.83	-0.07	-1.90	0.00	-0.03	0.25	5.47
	unobserved hhold hetero	0.46	9.84	0.06	1.66	0.18	6.96	0.00	0.00	2.26	59.71	1.32	18.72	0.13	2.84
	total hhold	**3.41**	**72.83**	**3.50**	**90.88**	**2.42**	**92.08**	**2.41**	**80.23**	**3.58**	**94.37**	**5.72**	**81.18**	**3.76**	**83.47**
Community determinants	health facilities	0.43	9.21	-0.02	-0.54	0.12	4.51	0.54	18.07	-0.07	-1.89	0.95	13.47	0.62	13.73
	public transport	0.09	1.89	0.30	7.89	0.08	3.06	0.04	1.46	0.19	4.92	0.01	0.09	0.04	0.92
	unobserved comm hetero	0.11	2.38	0.07	1.78	0.01	0.35	0.01	0.24	0.10	2.60	0.37	5.26	0.08	1.88
	total comm	**0.63**	**13.47**	**0.35**	**9.12**	**0.21**	**7.92**	**0.59**	**19.77**	**0.21**	**5.63**	**1.33**	**18.82**	**0.74**	**16.53**
Country determinants	**country effects**	**0.64**	**13.70**												
	total explained	**4.68**	**100.00**	**3.85**	**100.00**	**2.63**	**100.00**	**3.01**	**100.00**	**3.79**	**100.00**	**7.05**	**100.00**	**4.50**	**100.00**
	gap in IMR	**4.54**		**3.55**		**3.41**		**2.94**		**2.02**		**6.68**		**4.91**	

Table 7: Detailed decomposition results for the pooled sample and for each country separately.
NOTE: Absolute contributions and the gaps are expressed in percentage terms.

		MALI			BENIN			CAR		
		total	urban	rural	total	urban	rural	total	urban	rural
	Variables									
Proximate determinants	firstborn	**0.641*****	**0.325****	**0.711*****	0.131	0.247	0.119	0.737***	0.70***	0.777***
	birth order>4	0.180**	0.241	0.165**	0.382***	0.540*	0.350***	0.282**	0.054	0.388**
	mum's age at birth<20	0.395***	0.581***	0.360***	0.432***	0.625**	0.365***	0.315**	0.112	0.424**
	mum's age at birth>35	-0.114	0.014	-0.131	**-0.061**	**0.370**	**-0.211**	-0.048	-0.426	0.051
	short birth interval	0.738***	0.610***	0.763***	0.528***	0.424*	0.542***	0.878***	0.921***	0.860***
Socioeconomic determinants	mum not educated	0.331**	0.386**	0.119	0.281	0.004	0.370	0.084	0.071	0.199
	contraception	-0.198***	-0.286**	-0.158**	-0.016	-0.206	0.045	0.065	0.061	0.067
	age 1st marriage	-0.017*	-0.020	-0.015	**0.008**	**-0.036**	**0.021**	-0.008	0.024	-0.019
	sex of child	0.186***	0.011	0.220***	0.067	0.312*	-0.005	0.172*	0.271	0.128
	age of household head	0.001	0.007	0.000	0.002	-0.009	0.005	0.010	0.001	0.016***
	sex of household head	0.317***	0.105	0.380***	-0.021	-0.252	0.094	0.493**	0.298	0.795***
	toilet	-0.067	0.041	-0.076	-0.247*	-0.220	-0.248	-0.017	0.227	-0.105
	water	-0.024	-0.097	0.002	-0.109	-0.124	-0.085	**-0.240***	**-0.678*****	**-0.086**
	electricity	-0.115	-0.100	0.001	**-0.248**	**0.196**	**-0.468**	-0.198	-0.127	-0.175
	no finished floor	0.232**	0.186	0.267*	0.087	0.221	0.054	0.411**	0.249	0.510*
	asset1	0.083	0.359*	0.044	-0.016	0.361	-0.105	0.190	0.508*	0.011
	asset2	0.082	0.101	0.068	-0.020	0.310	-0.099	0.157	0.392*	0.018
Community determinants	health facilities	0.003	-0.196	0.001	-0.102	0.041	-0.144	-0.187	-0.367**	-0.069
	public transport	-0.0686	-0.000	-0.068	-0.061	0.022	-0.098	-0.055	0.033	-0.112
	constant	-2.993***	-2.915***	-2.872***	-3.344***	-2.688***	-3.629***	-3.989	-3.701***	-4.470***
	observations	20176	3535	16641	8124	2204	5920	5436	2039	3397
	variance of hh effect	0.301***	0.526**	0.261**	1.044***	1.617***	0.904***	0.000	0.000	0.000
	variance of comm effect	0.037***	0.000	0.042*	0.012	0.061	0.000	0.051	0.000	0.081
	mean prediction	0.152	0.123	0.159	0.102	0.078	0.112	0.096	0.078	0.108

		CHAD			NIGER			GUINEA		
		total	urban	rural	total	urban	rural	total	urban	rural
Proximate determinants	firstborn	0.436***	0.155	0.521***	0.215*	-0.036	0.248**	0.169	0.246	0.146
	birth order>4	-0.109	0.026	-0.170	**-0.036**	**0.400****	**-0.097**	0.269***	0.239	0.288***
	mum's age at birth<20	**0.176**	**0.599*****	**0.047**	0.392***	0.529**	0.375***	0.285**	-0.118	0.380***
	mum's age at birth>35	-0.123	-0.427	-0.045	-0.192	0.130	-0.209	0.089	0.139	0.071
	short birth interval	0.756***	0.619***	0.781***	**0.317*****	**0.635*****	**0.277*****	0.356***	0.623***	0.286***
Socioeconomic determinants	mum not educated	**0.320**	**0.171**	**1.437****	0.290	0.147	0.295	0.335**	0.302	0.438
	contraception	-0.129	-0.272	0.044	-0.533***	-0.749***	-0.400**	-0.315**	-0.274	-0.310**
	age 1st marriage	-0.019	0.022	-0.029	-0.012	0.014	-0.015	0.0234**	0.016	0.024*
	sex of child	0.095	0.006	0.102	0.120	-0.160	0.158*	0.200***	0.357***	0.163**
	age of household head	-0.001	0.006	-0.003	0.005	0.004	0.004	0.003	0.010	0.002
	sex of household head	0.083	-0.11	0.213	0.367**	-0.016	0.424***	-0.115	-0.215	-0.056
	toilet	0.246*	0.045	0.272	**0.167**	**0.590****	**0.011**	-0.092	-0.159	-0.084
	water	**-0.435*****	**-0.221***	**-0.540*****	**-0.217***	**-0.708*****	**-0.157**	-0.244*	-0.009	-0.503
	electricity	-0.012	-0.095	0.531	-0.842***	-0.605***		**-0.142**	**0.077**	**-1.254****
	no finished floor	0.451*	0.330	1.148	**0.063**	**0.392***	**-0.177**	0.092	0.375	0.017
	asset1	-0.019	-0.095	-0.027	0.023	0.052	0.021	0.154	0.258	0.096
	asset2	-0.079	0.018	-0.104	-0.076	-0.219	-0.062	0.081	0.341	0.014
Community determinants	health facilities	0.018	0.002	0.123	-0.234*	-0.282*	-0.278	-0.291***	-0.335**	-0.260**
	public transport	-0.066	0.150	-0.667	-0.002	0.343	-0.016	0.059	-0.378*	0.102
	constant	-3.373***	-3.383***	-5.242***	-3.051***	-2.944***	-2.868***	-3.191***	-3.468***	-3.214***
	observations	8738	3728	5010	10596	2480	8116	9403	2410	6993
	variance of hh effect	1.671***	0.000	2.638***	1.329***	0.321*	1.478***	0.496***	0.252	0.540***
	variance of comm effect	0.117***	0.000	0.181**	0.123***	0.000	0.149***	0.047*	0.000	0.057*
	mean prediction	0.120	0.104	0.124	0.150	0.094	0.160	0.128	0.090	0.139

Table 8: Regression coefficients for the total, urban and rural sample.
NOTE: Significance at: * 10%, ** 5%, * 1%.**
Coefficients in bold indicate that they differ significantly between urban and rural subsamples at the 10% level.
For rural Niger, there is no coefficient for electricity, because of too little variation.

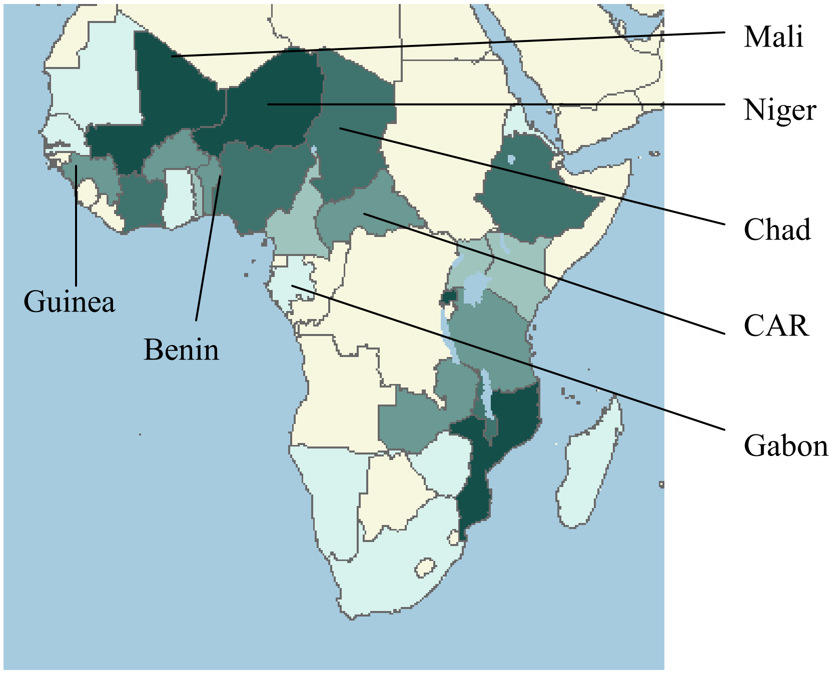

Figure 1: Clustering of infant mortality in Sub-Saharan Africa.
NOTE: Darker colors indicate higher infant mortality rates. Countries in white indicate no data was available.
Source: Demographic and Health Survey StatMapper

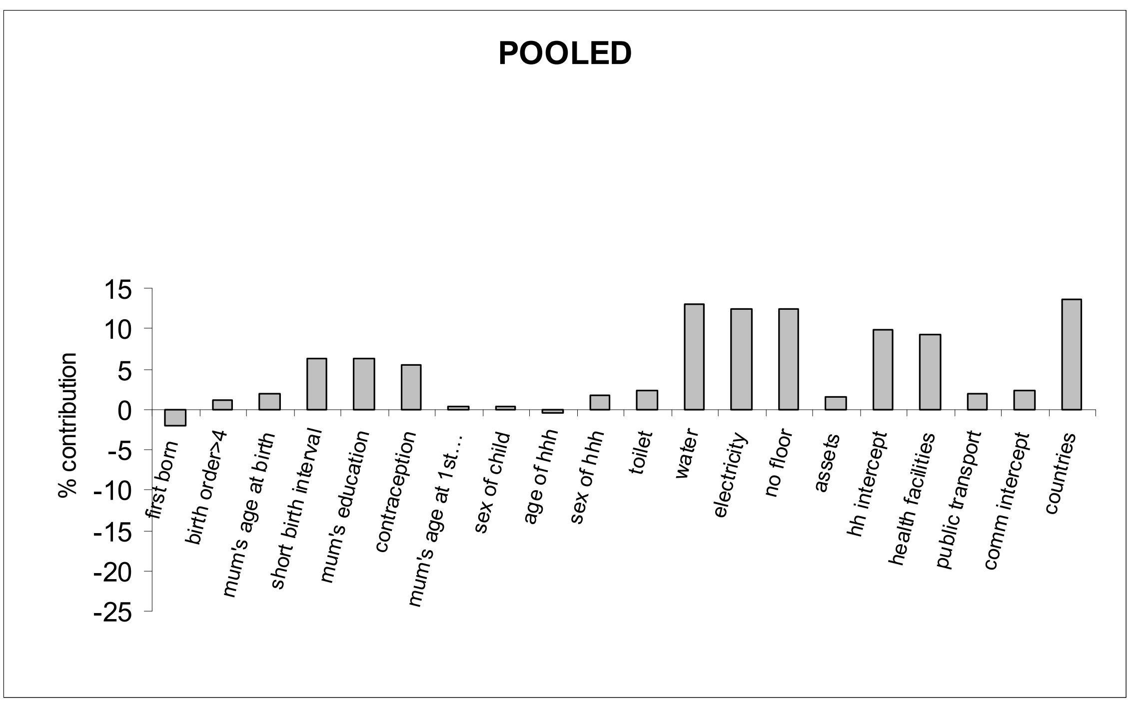

Figure 2: Detailed decomposition results on the sample pooled across countries.

NOTES

1 Calculated using data from the World Development Indicators (World Bank 2006) and DHS (Statcompiler).

2 Gabon's GNP per capita is about 5 times the average of the other six countries included in the study and four times the average for Sub-Saharan Africa (World Bank 2006). This is in large part due to offshore oil production.

3 We could not include information on breastfeeding practices because DHS only contain the relevant data for the 5 lastborn children. Breastfeeding may be considered endogenous in stunting/mortality regressions. Further, we chose not to include an indicator of mother's nutritional status because current nutritional status might not be a good indicator of health at the time of pregnancy and because a considerable number of women were pregnant at time of survey.

4 We did experiment with including the size of the households and ratio of number of persons per room but household size appeared very much related to survival and gave counterintuitive results.

5 Using such a list of assets for both urban and rural areas from a common set of assets may understate the wealth of rural households because the DHS generally contain more information on assets that are more common to urban areas (eg. fridge, television). Households in rural areas may have a range of resources that are often not recorded in DHS, like land, rights to fishing, gathering or grazing, or the space and resources to keep animals. It might also be that the correlation between certain assets and wealth differs between urban and rural areas, although Menon, Ruel & Morris (2000) have found no clear evidence of this.

6 We also tried including other community variables such as the existence of a market place, but this showed no effect. Further, we experimented with creating an index of public services that combines information on existence of a shop, public transport, market, post, bank, and garbage collection in the community. However, these services were not consistently available for all countries and were not significant in country specific models. For some countries the data contain more detailed information on health services but proximity is the only information that is available across the entire set of countries.

7 When decomposing rural-urban gaps in infant mortality into gaps in the determinants, it is important to have sufficient common support of the determinants across urban/rural areas. If not, a covariate might be just picking up the rural-urban disparity, or might be capturing an 'outlier' effect. In this respect,

Table 5 shows that almost all rural women are non-educated. We experimented with considering 'incomplete primary education' also as valid education, but then we did not find any effect of maternal education on infant mortality. This is probably because 'incomplete' education might mean different things, and therefore has no clear effect on infant mortality. A further issue might be the very low average electricity access in rural areas. However, we redid the entire analysis (also the country specific regressions) excluding the electricity variable, and found that the effect of the other variables remained unchanged.

8 Neglecting unobserved heterogeneity in non-linear models causes coefficients to be inconsistent, although consistency of the average partial effects is preserved (Wooldridge 2002).

9 We considered using duration analysis to model child survival. However, an Oaxaca-type decomposition has not been developed for these kind of models and is complicated by rural-urban differences in duration dependence. Furthermore, duration models would not add much compared to binary models in this particular application since there are no time-varying covariates.

10 Strictly speaking the random intercepts are parameters to be estimated and so one logic would place them with the contribution of the difference in the coefficients in the decomposition. We prefer to place them with the covariate contribution since they essentially reflect differences in the distributions of determinants, albeit unobservable ones.

11 Several weighting alternatives have been suggested in the decomposition literature (see e.g. Neumark 1988; Oaxaca & Ransom 1994). Using the pooled coefficients as weighting factors for differences in the distribution of the covariates seems most justified in our case since neither the rural nor the urban model can be interpreted as the natural order from which the other deviates due to discriminatory behavior .

12 This decomposition is an approximation. The probability of dying is modeled as a non-linear logistic function over the distribution of the household and community intercepts. In the decomposition we estimate this probability as a logistic function evaluated at the posterior mean of these household and community intercepts, which, because of the non-linearity, is not the same as the former.

13 Unlike in the linear case, the independent contribution of a covariate depends on the values of the other covariates. This implies that the order of switching the distributions could affect the estimated contribution of each covariate. To check sensitivity, we experimented with randomizing the order of the switching of covariates as suggested by Fairlie (2005) and found that the results were very robust.

14 Since we use sampling with replacement, some rural children may be more than once in the subsample that is used for the matching. The order of these 'duplicate' children is then randomized to match them with an urban child.

15 Increasing the number of replications further did not change decomposition results significantly.

16 It must be noted that when pooling across countries, the data is in fact organized on 4 levels: children, households, communities and countries. We chose to include fixed as opposed to random effects to capture country-specific characteristics. Because we only have 6 countries, fixed effects are straightforward to estimate and do not require the assumption of independence of the other covariates.

17 We also tried including sex of child interacted with being the first born, and found that there also is a male disadvantage for firstborn children. The male disadvantage in child survival has commonly been found in other studies (see e.g. Curtis *et al* 1993; Sastry 1996, 1997b; Bolstad & Manda 2001).

18 Using under-five instead of infant mortality increases the proportion of deaths and makes the unobserved components easier to identify. However to still have sufficient observations, it is required to extend the time period in which births took place (we used 15-5 years before the survey) and therefore it less likely that current household conditions reflect those within the first years of life.

19 In these detailed decomposition results, the percentage of the gap that is explained does not exactly equal the 100 percent mentioned before. This is due to the approximation in the contribution of the unobservables mentioned in Endnote 12. The same holds for the country specific results.