Cause-specific neonatal deaths in rural Bangladesh, 1987-2005: Levels, trends and determinants

Unnati Rani Saha¹, Arthur van Soest² and Govert E. Bijwaard³

¹Tilburg University, Warandelaan 2, 5037 AB, 5000 LE Tilburg, The Netherlands; ICDDR,B, Dhaka, Bangladesh; Emails: unnati.saha01@gmail.com

²Tilburg University, Warandelaan 2, 5037 AB, 5000 LE Tilburg, The Netherlands; Email: avas@uvt.nl

³Netherlands Interdisciplinary Demographic Institute (NIDI); Email: bijwaard@nidi.nl

Running title: Cause-specific neonatal deaths

We wish to thank three anonymous reviewers, the editors, Katie Carman, Sabu Padmadas and Frederic Vermeulen for their valuable comments which helped to improve the paper. We are grateful to Carel van Mels for providing us with the verbal autopsy codes used in Matlab and to Nurul Alam and Mahfuzur Rahman for the clarification of verbal autopsy data.

This paper reports on an analysis of neonatal mortality from communicable and non-communicable diseases in Bangladesh. The competing-risks model employed incorporated both observed and unobserved heterogeneity and allowed the two heterogeneity terms to be correlated. The data used came from the Health and Demographic Surveillance System, Matlab. The results confirm the conclusions of previous studies about the levels, trends, and causes of neonatal death in the Matlab area: the education of the mother helps protect her children from death from both communicable and non-communicable diseases; the children of a father in a low-status occupation are particularly vulnerable to death from communicable diseases, and children born to mothers aged less than 20 face a particularly high risk of dying from a non-communicable disease. The risks of dying from a communicable disease and from a non-communicable disease were both found to fall significantly as the distance to the nearest health centre decreased.

Keywords: millennium development goals; neonatal deaths; competing risks models; unobserved heterogeneity; Bangladesh

1. Introduction

The achievement of the fourth Millennium Development Goal, which aims to reduce mortality amongst children under the age of 5 to one third of its 1990 level by the year 2015, remains one of the United Nations' global priorities (United Nations 2001). Recent trends in mortality suggest, however, that unless there are substantial reductions in neonatal mortality, that Goal will not be achieved (Lawn et al. 2005). Reviews of the global situation suggest that further reductions in mortality will only be possible if interventions are put in place to address the current shortcomings in the healthcare of mothers, newborn infants, and young children (Jones et al. 2003; Black et al. 2010). Those implementing such interventions rely on cause-of-death information in order to

prioritize disease-specific interventions and to increase their effectiveness (Baqui et al. 2001; Lawn et al. 2006).

The reduction of neonatal mortality is a particularly important and pressing issue in Bangladesh. Although infant and child mortality rates in Bangladesh declined sharply during the last decades of the twentieth century, the rate of decline has slowed down in recent years, particularly amongst neonates. Deaths occurring in the first month of life represent an increasing proportion of all deaths under age 5 in Bangladesh (NIPORT et al. 2011). Recent estimates suggest that in 2009, 57 per cent of deaths in this age range occurred during the neonatal period in that country, whereas the figure had been only 39 per cent two decades earlier (Liu et al. 2011). Here we report on an analysis of the levels of, and trends in, neonatal deaths due to communicable and non-communicable diseases in Bangladesh and the associated risk factors, both observed and unobserved. We believe the findings will help to inform the design of policies intended to reduce neonatal mortality in Bangladesh in particular, but also in many other countries in the developing world.

The study was focused on two classes of cause of death—communicable and non-communicable diseases. We estimated competing-risk duration models of causes of death amongst infants during the neonatal period, that is, during the first 28 days after birth. Children who survived the neonatal period were treated as censored observations. The modelling procedure, which was more flexible than in many previous studies of the determinants of cause of death, combined a piecewise constant baseline hazard with proportionality assumptions about the influence of observed and unobserved risk factors for each cause. The model allowed the unobserved heterogeneity components in the two competing hazards for deaths due to communicable and non-communicable diseases to be correlated.

The estimations were based on prospective panel data from the Matlab region in Bangladesh, where the health of mothers and their children was monitored from 1987 until 2005. The data were drawn from the villages in the Matlab region, separated into two areas (see ICDDR,B 2006 for details). In the first area of 70 adjacent villages, an intervention to reduce neonatal mortality was already in place with non-standard health services being provided by the International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B); this we will refer to as the 'ICDDR,B area'. The remaining area, which we will refer to as the 'comparison area', had only standard government-provided healthcare facilities. By comparing the two areas we hoped to gain insight into how the different healthcare services available shaped child health epidemiology within the two populations over the study period.

2. Background

The Health and Demographic Surveillance System (HDSS) run by ICDDR,B routinely records all births and deaths for a total population of about 220,000 (ICDDR,B 2006). Causes of death are recorded through verbal autopsy. The HDSS also incorporates several indicators of the socioeconomic status of each household. The HDSS thus plays an important role by providing accurate information on vital events, including details of cause of death which are often unavailable in many data sets where resources for data collection are more constrained.

In 1977, ICDDR,B started to provide extensive additional maternal health, child health, and family planning services, over and above the existing government health services, in half of the region covered by the HDSS. This was our 'ICDDR,B area'. The other half of the region covered by the HDSS, continued to receive only standard government health services, and this was our 'comparison area'. Both areas were typical of rural Bangladesh. The programme of maternal and child health in the ICDDR,B area was introduced incrementally, phase by phase, (Phillips et al.

1984) and included the provision of domiciliary family planning services, simple education on nutrition, tetanus toxoid immunization for women of reproductive age, measles immunization, and community-based oral rehydration therapy. In addition several sub-centres provided treatment for minor illnesses and basic emergency obstetric care, and a permanent hospital provided treatment for diarrhoeal diseases. In order to understand how these better health services shaped child health, we compared the data from the ICDDR,B area with that from the comparison area,

The HDSS data have been used in many previous studies of maternal and child health and mortality, such as those of Hale et al. (2006), DaVanzo et al. (2008) and Saha and van Soest (2011). Hale et al. found that about 20 per cent of the difference between the ICDDR,B area and the comparison area in rates of infant and child mortality could be explained by differences in the patterns of reproduction in the two areas. DaVanzo et al. and Saha and van Soest analysed the relationship between the mortality outcomes of successive children and their birth spacing. These studies, and many others, did not take cause of death into account.

Studies of mortality using the HDSS data which do take cause of death into account include those of Bhatia (1989), Fauveau et al. (1991), Karar et al. (2009), and Chowdhury et al. (2010). Other studies have analysed causes of child mortality throughout Bangladesh using the Bangladesh Demographic and Health Surveys; see, for example, Baqui et al. (2006) and Liu et al. (2011). Most studies of cause-specific neonatal deaths in developing countries, including those in the Matlab region, report neonatal mortality rates by age-group within the first month of life, because the age at which a death occurs can offer insight into its cause (Chen et al. 1980; Bhatia 1989; Baqui et al. 2001; Lawn et al. 2006; Karar et al. 2009; Chowdhury et al. 2010). From the trends in mortality observed, however, it has become apparent that factors other than the health and family planning interventions put in place in Bangladesh influence levels of mortality within the country (Bhatia 1989). In order to target the interventions more effectively it therefore seems

important to discover which underlying factors are associated with particular causes of death, and to do so was the aim of the current study.

Duration models are often used in demographic and epidemiological research when the events to be modelled are associated with periods of time, such as ‘time until marriage’, ‘time from marriage until the birth of a first child’, or ‘time until death’ (Cox 1959; Heckman et al. 1985). Analysis of ‘time until death’ for alternative, or ‘competing’, causes to which a person might succumb yields knowledge about the timing of death from various types of disease, and such knowledge helps make it possible to target prevention policies more effectively. The analysis of cause-specific death can be related to the concept of competing risks (Cornfield 1957; Fine and Gray 1999; Coviello and Boggess 2004). In a ‘competing risk’ situation in which an individual is at risk of death from several causes, the analysis of any one specific cause of death has to take account of the risk of death from all the ‘competing’ causes. For example, in one study it was found that the probability of a female developing cancer at some point during her life had increased by 25 per cent over a seven-year period, but that the largest part of the increase was accounted for by decreases in other causes of mortality (Goldberg et al. 1956). To our knowledge, our study is the first to apply a competing-risk model to the study of neonatal mortality in Bangladesh.

Studies of child deaths have shown that the survival chances of the children within a particular family are correlated and that variation in the risk of death between families remains even after controlling for observed covariates such as the age of the mother at the time of a birth, the sex of the child, or its race. This variation can be attributed to unobserved mother-specific heterogeneity that leads to risk factors that are common to all children of the same mother. For example, the children from a particular mother may all suffer from the same adverse genetic traits, or they may all suffer from the inability of their parents to take good care of them, or they may all be exposed to the same environmental factors (Mosley and Chen 1984; DasGupta 1990;

Arulampalam and Bhalotra 2006). Such common risk factors will often be specific to certain types of risks, such as communicable or non-communicable diseases. If, for example, a mother has a high propensity to give birth to children of low birth weight (a non-communicable cause of death), it is possible that all births to this mother will show evidence of low birth weight. Communicable diseases have risks in common within the same family since children may catch infectious diseases from their siblings (see, for example, DaVanzo et al. 2008) or they may all be put at risk by local environmental factors such as an unsafe water supply or limited access to healthcare.

3. Data

We combined the HDSS data on all live births and deaths of children for the ICDDR,B and comparison areas obtained between 1987 and 2005. The dataset contained the records of 107,367 singleton live births, of which 57,830 were reported in the comparison area and 49,837 in the ICDDR,B area. The data included reports of 4,047 neonatal deaths with their underlying causes, 2,446 in the comparison area and 1,601 in the ICDDR,B area. Additional data on the education of both the mother and father, on the occupation of the father, and on the source of the family's drinking water were obtained from the 1982, 1996 and 2005 socio-economic census data that are recorded in HDSS.

3.1. Assignment of cause of death

The verbal autopsy method used by HDSS to assign the cause of an infant's death has been described extensively in the literature (see, for example, Chowdhury et al, 2010 and Karar et al. 2009), but it may be helpful to summarize here how a verbal autopsy is undertaken. There are several steps in the process. All deaths in Matlab are recorded during the monthly house-to-house visits. Shortly after recording the death of a child, a one-page Death Form is completed by an experienced, but medically untrained, interviewer from data supplied by the mother of the child, or

in her absence, by a close relative or neighbour. This form contains information on the signs and symptoms leading to death and medical consultation prior to death. The Death Form used in Matlab and the selection and training of interviewers were all revised in 2003 to meet WHO standards (see Karar et al. 2009, Table 1). Finally, a medical assistant, with three years training in medicine and experience in verbal autopsy techniques, reviews the Death Forms and, using his own judgement, assigns the direct and underlying causes of death. The medical assistant worked at the ICDDR,B hospital in Matlab; the same medical assistant undertook this task from 1986 until 2006 in both the ICDDR,B and the comparison areas. If necessary, the family of the deceased was visited to collect additional information. A code was then assigned to the cause of death, selected from a list of 97 possible codes derived from the ‘basic tabulation list’ drawn up by the World Health Organization International Statistical Classification of Disease, Injuries and Causes of Death (ICD-9 and ICD-10). The assignment of cause of death using verbal autopsy or vital registration data of WHO is described in detail in the literature (Fauveau et al. 1994; Baqui et al. 2001; Adjuik et al. 2006; Lawn et al. 2006; Karar et al. 2009; Chowdhury et al. 2010; Ronsmans et al. 2010).

The method of verbal autopsy has been widely used in many developing countries; see, e.g., Soleman et al. (2006) or Garenne and Fauvau (2005). Validation studies have shown that the method is more reliable for some causes of death than for others (e.g., Marsh et al. 2003; Freeman et al. 2005), but since we grouped causes of death into two broad categories, communicable and non-communicable diseases, our study was less vulnerable than it would have otherwise to the misclassification of a cause. Moreover, Chowdhury et al. (2010) describe a study carried out in Matlab in which three physicians independently reviewed verbal autopsy interview data before a cause of death was assigned and entered on a death certificate. The prevalence rates for deaths due to communicable and non-communicable diseases in the period 2003-2004 in our data are similar to

those in Chowdhury et al., suggesting that the data on the distinction between communicable and non-communicable causes of death based upon our verbal autopsy data are of good quality..

Once the cause of death had been identified, we followed Karar et al. (2009) and placed each individual cause of neonatal death listed in Table 1 into one of 2 classes: ‘communicable’ or ‘non-communicable’ diseases. Acute respiratory infections, diarrhoea, dysentery, sepsis, meningitis, hepatitis, chickenpox, neonatal tetanus, as well as diphtheria, whooping cough, measles, and acute polio (covered by the Extended Program on Immunization) were placed in the ‘communicable diseases’ class, along with deaths from other viral diseases. In the non-communicable class, we placed deaths of children who died after being born pre-term or with a low birth weight, those who died from conditions related to their gestation or birth—such as birth asphyxia, birth trauma or cord haemorrhage, other obstetric complications of the newborn, or congenital abnormality— and those who died from neonatal infections, sudden infant death, and unspecified or miscellaneous other neonatal deaths. The additional four categories in the right hand column of Table 1 will be discussed below. The coding and classification was done using the International Classification of Diseases, Revision 9 (ICD9) and Revision 10 (ICD10) (see <http://apps.who.int/classifications/icd10/browse/2010/en#/I> and Table 1). Matlab HDSS assigned the cause of death following ICD-9 code for the period of 1986-2002 and 2003 onwards they used ICD-10 codes to assign the cause of death. Since our study period covers 1987-2005, we used both ICD9 and ICD10.

Table 1 about here

3.2 Socio-demographic variables

When selecting variables that drive mortality as covariates in our models, we followed the literature, and in particular Hale et al. (2006) and Saha and van Soest (2011), who used the same data sources as we did. The existing studies suggest that the likelihood of a child surviving his or her first year varies with the sex, birth order, and year of birth of the child in question. By comparing across the birth cohorts observed between 1987 and 2005 it has been possible to capture the time trends in mortality and the epidemiological shifts that occurred in each of the three decades concerned. We used two dummies, dividing the total time covered by our data into the periods before 1993, 1993 to 1999, and 2000 to 2005. Three birth-order variables were included to capture sibling effects on the risks of cause specific mortality; birth order was divided into ‘first-order births’, ‘second- and third-order births’ and ‘forth-order births or above’. The following characteristics of parents thought to have a potential role in child mortality were also included in our models: the religion of the family, or at least whether or not they were Muslim; level of education received by both parents; occupation of the father, day labourers being differentiated from the rest; and the mother’s age at the birth of each of her children. Mother’s education was taken to be a proxy for child care skills and the ability to use modern healthcare services. Both father’s education and occupation were included as indicators of socioeconomic status. Mortality also depends on the characteristics of the community in which individual families live, such as whether or not residents receive their water via a pipe, and the distance to the nearest health facility. The *distance to health centre* variable captures the availability of health services and the *source of drinking water* variable represents environmental effects on a child’s survival chances. Summary statistics of the chosen variables, comparing those for the ICDDR,B area with those for the comparisons area are presented in Table 2.

Table 2 about here

3.3. The distribution of neonatal deaths by cause

Table 3 shows cause-specific mortality rates per 1,000 live births during the neonatal period of 28 days in both the ICDDR,B and the comparison areas for the period 1987-2005. During the period 1987-2005, a neonatal mortality rate of 32.3 per 1,000 live births was recorded in the ICDDR,B area, where high-quality healthcare services were available; in the comparison area, where only standard government health services were in place, a rate of 42.3 was recorded. The latter figure is close to Bangladesh's national neonatal mortality rate of about 41 per 1,000 live births (see NIPORT et al. 2005). The largest single cause of neonatal deaths was non-communicable diseases. They accounted for 28.1 and 33.0 deaths per 1,000 live births in the ICDDR,B and comparison areas, respectively. Prematurity, or low birth weight, was the leading cause of death listed in Table 3, accounting for 14.9 in deaths per 1,000 live births in the ICDDR,B area, and 15.5 in the comparison area. The second most important cause-of-death category in Table 3 is 'diagnosis not possible'. (A sensitivity analysis in which this group was excluded completely from the analysis, rather than being included in the non-communicable diseases category, gave conclusions similar to those of our main analysis; details are available upon request.) Among the deaths attributed to communicable diseases, the acute respiratory infections accounted for the greatest number: about 3.5 and 6.5 deaths per 1,000 live births in the ICDDR,B and comparison areas, respectively.

Table 3 about here

Before considering our competing-risks model we present the non-parametric cumulative incidence functions for the different causes of death, to show how the mortality rate from communicable and non-communicable diseases changes with the age of the child. The value of the *cumulative incidence function* of cause j at time t indicates the probability that a child will die from

cause j before reaching age t , and is a function of the hazards of all the competing events (see, for example, Coviello and Boggess 2004). It was estimated using sample fractions instead of probabilities: for cause j the estimated value of the cumulative incidence function at t days of age is the ratio of the number of children who died from that cause at, or before, age t divided by the total number of children in the sample. Figure 1 shows the cumulative incidence functions for neonatal deaths attributed to communicable and non-communicable diseases derived from the complete sample for each of the two areas. The Figure shows that most deaths from non-communicable diseases occur in the first few days after birth, whereas deaths from communicable diseases are more evenly spread over the whole of the neonatal period. The shape of the curves are quite similar in the two areas, although the levels of mortality are not.

--- **Figure 1 approximately here** ---

4. Model

The modelling approach we adopted in our analysis was built on the concept of competing risks. We take δ_{ij} to be equal to 1 if child i dies from cause j ($j=1,2$) during the first 28 days after birth. We assumed that each neonatal death was the result of just one single cause, the main cause of death, so we exclude the possibility that the child dies from multiple causes simultaneously. In our model we use two competing causes of death: a child can die from a communicable disease or a non-communicable disease. The hazard of dying from cause j at age t is denoted by $\lambda_j(t)$, where t refers to the age of the child in days and can range in value from 0 to 27. In the competing- risks model, a child is observed until it dies from one of the two possible causes of death or until it survives to the end of the neonatal period. The observed survival time, or age at death, of child i is

therefore given by $T_i = \min(T_{i1}, T_{i2}, C_i)$. In our case $C_i = 28$ days. So T_i is the age of death in days in the case of a neonatal death and 28 days when a child survives the neonatal period.

For our model, $\lambda_1(t)$ is the hazard of dying from a communicable disease and $\lambda_2(t)$ the hazard of dying from a non-communicable disease at age t . The hazard of dying at age t from all possible causes is given in our model by $\lambda_1(t) + \lambda_2(t)$. This sum corresponds to the single hazard of dying in the basic hazard model.

The hazard rates are specified as the following mixed proportional hazards (see, for example, Manton et al. 1981):

$$\lambda_j(t | x, v_j) = \lambda_{0j}(t) \exp(x\beta_j)v_j, j = 1, 2 \quad (1)$$

where $\exp(z)$ denotes e^z . The hazard rates are functions of age t , the explanatory variables x ; selected characteristics of the child, the mother and the community in which they live, which in our model are assumed not to vary over the neonatal period; and mother-specific (unobserved) heterogeneity that remains constant over the time period that the mother has her children, such as generic traits, parental care abilities, or environmental factors v_j . The relative risks associated with explanatory variables x are captured by $\exp(x\beta_j)$, $j=1,2$. Age dependence of the hazard is incorporated using a piecewise constant baseline hazard $\lambda_{0j}(t)$. The baseline hazard captures the age-specific shape of the hazard, which is common to all individuals. Thus for each cause of death j , we have

$$\lambda_{0j}(t) = \exp(\beta_{0j}) \sum_{h=1}^H \exp(\gamma_{hj}) I_h(t) \text{ with } I_h(t) = I[t_{h-1} \leq t < t_h]$$

In the above equation H is the number of intervals in which the complete neonatal period of 28 days is divided, These intervals are indexed using $h=1, \dots, H$, $I[t_{h-1} \leq t < t_h]$ is the indicator function (1 inside the interval $[t_{h-1}, t_h]$, 0 outside the interval $[t_{h-1}, t_h]$, $t_0 = 0$, and $t_H = 28$ days. We experimented with several partitions of the 28 days into age intervals and found the model

performed best when $H=5$ and the age intervals 0 (dies before end of first day), 1 (dies on second day), 2, 3-6 days and 7-27 days (survives the first week but dies before the end of the first four weeks) were used. For identification of all the parameters of the baseline hazard we need to restrict one of the γ_{hj} ($h=1, \dots, H$) to 0 for each j . We chose $\gamma_{Hj} = 0$. Thus, β_{0j} determines the hazard of death in the final age interval and the other γ_{hj} terms determine the ratio of the hazard in each previous interval to the hazard in the final interval.

We wished to emphasise the specification of unobserved heterogeneity v_j , $j=1, 2$, in order to capture unobserved factors which affected the chances that a child would survive its first month of life. Had these factors been ignored our parameter estimates might have been biased. In principle, unobserved heterogeneity can be child-specific, mother-specific, or community-specific. Following several previous studies which emphasized the role of mother-specific heterogeneity or ‘frailty,’ (Sastry 1997; Arulampalam and Bhalotra 2006) we focused on mother-specific heterogeneity; but community-specific unobserved heterogeneity could be incorporated in a very similar way and is discussed briefly in Section 5.1. The unobserved heterogeneity terms, $v_j > 0$ ($j=1, 2$), are time-independent and also independent of the observed characteristics, x , but they were allowed to be correlated with each other, thus inducing potential correlation between the hazards of dying from different causes of death. The correlation was easy to incorporate using the following discrete distribution of $v = (v_1, v_2)$:

$$\begin{aligned} P(\ln v_1 = -\alpha_{11}, \ln v_2 = -\alpha_{21} - \alpha_{22}) &= p_1 p_2; \\ P(\ln v_1 = -\alpha_{11}, \ln v_2 = -\alpha_{21} + \alpha_{22}) &= p_1 (1 - p_2); \\ P(\ln v_1 = \alpha_{11}, \ln v_2 = \alpha_{21} - \alpha_{22}) &= (1 - p_1) p_2; \\ P(\ln v_1 = \alpha_{11}, \ln v_2 = \alpha_{21} + \alpha_{22}) &= (1 - p_1) (1 - p_2). \end{aligned}$$

Here p_1 , p_2 , α_{11} , α_{21} , and α_{22} are parameters to be estimated (with p_1 , p_2 between 0 and 1). The two hazards are correlated if and only if $\alpha_{21} \neq 0$. These parameters were estimated jointly using the

maximum likelihood estimation technique, and details of the likelihood function are available upon request. The covariance matrix of the unobserved heterogeneity terms was estimated *ex post*, since it is a function of the model parameters. This also applies to the *total survival* and *cumulative incidence functions*. The *total survival function*, at age t is defined as the probability of not dying from either a communicable or a non-communicable disease before age t (conditional on observed and unobserved heterogeneity). It is given by:

$$S(t | x, v_1, v_2) = \Pr(T \geq t | x, v_1, v_2) = \exp\left(-\sum_{j=1}^2 v_j \int_0^t \lambda_{0j}(s) \exp(x\beta_j) ds\right) \quad (3)$$

The *cumulative incidence function* of cause j is the probability of dying due to cause j before age t . In Section 3, we have already presented the empirical cumulative incidence functions for various causes for the complete samples from the two areas. Based upon the model, we were also able to estimate the cumulative incidence functions for a specific child, that is, conditional on observed and unobserved heterogeneity. They are given by:

$$F_j(t | x, v_1, v_2) = \int_0^t v_j \lambda_{0j}(s) \exp(x\beta_j) S(s | x, v_1, v_2) ds \quad (4)$$

Integrating out the observed and unobserved heterogeneity, we were also able to obtain the average total survival and cumulative incidence functions for a child with given observed individual and family characteristics x and with mother specific unobserved heterogeneity $v = (v_1, v_2)$. Note that the sum of all cumulative incidence functions at a given age is equal to 1 minus the total survival function at that age, i.e., $\sum_{j=1}^2 F_j(t | x, v_1, v_2) = 1 - S(t | x, v_1, v_2)$.

5. Results

5.1 Model selection

We estimated several specifications of our model, with and without several forms of unobserved heterogeneity, and for the ICDDR,B area and comparison area separately as well as jointly, either

with an area dummy in each hazard or with interactions between the area dummy and all the covariates. Table 4 summarizes the goodness of fit of all these specifications. The first double column presents the results of the joint models without interactions, and the second double column presents the results of the joint models with interactions in the proportional hazard part. The third double column presents the results of separate models for the two areas, which is equivalent to including interactions in the baseline hazards, the unobserved heterogeneity parameters, and the proportional hazard part. The rows in Table 4 represent different versions of the model with different specifications of unobserved heterogeneity. Row 1 represents a version where there is assumed to be no unobserved heterogeneity; row 2 a version with discrete mother-specific, unobserved heterogeneity as specified in Section 4; row 3 a version incorporating discrete, community-specific, unobserved heterogeneity; row 4 a version including bivariate normal, mother-specific, unobserved heterogeneity.

Table 4 about here

The maximum values of the likelihood function for alternative model specifications can be compared across the columns of Table 4 using standard likelihood ratio tests (see, e.g. Cameron & Trivedi 2005). The conclusions are the same for each row of the table: the joint model with only an area dummy (double column 1) cannot be rejected when compared against either of the two more general models that allowed for more differences between the coefficients for the ICDDR,B area and the comparison area. This suggests we can focus on the more parsimonious models in column 1.

Comparisons between the rows of Table 4 is less straightforward. The models in the first row are nested within the corresponding models shown in the other rows but, since the models without unobserved heterogeneity fix the parameter values to the lower limits of the parameter range, standard likelihood ratio tests cannot be used (see, for example, Self and Liang 1987).

Nevertheless, according to common selection criteria for econometric models, the models in row 1 of Table 4 are clearly outperformed by their counterparts in row 2, which include discrete, mother-specific, unobserved heterogeneity. By 'common selection criteria' we refer to such criteria as the Akaike Information Criterion, according to which the lower the value of the calculation $2(\text{number of parameters} - \log \text{likelihood})$, the better the fit of the model, or the Bayes Information Criterion, where again a lower value of the calculation $(-2 \log \text{likelihood} + (\text{number of parameters})\log(\text{number of observations}))$ indicates a better fit). Applying the same criteria to compare the non-nested models, the models in row 2 also outperform the models in row 3, which include community-specific unobserved heterogeneity, and the models in row 4, which include bivariate normal unobserved heterogeneity. The bivariate normal models have fewer parameters: the unobserved heterogeneity terms are the exponentials of two terms with a bivariate normal distribution with mean 0 and an arbitrary covariance matrix specified by three parameters, in contrast to the five parameters included in the discrete distribution models shown in row 2. But the smaller number of parameters of the bivariate normal model does not compensate for the difference between the log likelihoods in row 4 and row 2. For example, calculating the Bayes Information Criterion, since it has a larger penalty than the Akaike Information Criterion when more parameters are included in a model, taking the first column and comparing the second and fourth rows yield figures of 58,594 and 58,604, respectively. The smaller figures for row 2 indicate that we should prefer the models from this row, which incorporate discrete, mother-specific, unobserved heterogeneity. The way unobserved heterogeneity is captured in the models in row 2 thus becomes our benchmark specification of unobserved heterogeneity.

5.2 Parameter estimates

The parameter estimates for certain models, selected from amongst those in Table 4, are presented in Tables 5 and 6. Table 5 presents, as hazard ratios, $\exp(\beta_j)$, the parameter estimates for the joint model, first without unobserved heterogeneity (column 1, row 1 in Table 4), and then with discrete, mother-specific, unobserved heterogeneity (column 1, row 2 in Table 4), which is our benchmark specification. While the latter is our preferred model, we also present the model without unobserved heterogeneity, which we used to analyse the effect of incorporating unobserved heterogeneity on the other parameter estimates. In fact, the effect of incorporating unobserved heterogeneity is quite small: the parameter estimates derived from the two specifications, with or without unobserved heterogeneity, are very similar, as are their significance levels. This similarity is somewhat surprising because the model with unobserved heterogeneity fits the data much better and the estimation results of the model with unobserved heterogeneity also imply a significant correlation between the two unobserved heterogeneity terms and the hazard of dying from either a communicable disease or a non-communicable disease. The estimate of the auxiliary parameter, α_{21} , which drives this correlation, is not presented in Table 5 but is significant. The estimates of the covariance structure imply a correlation of 0.456 between the frailty terms $\ln(v_1)$ and $\ln(v_2)$ in the two hazards. The estimated variances for $\ln(v_1)$ and $\ln(v_2)$ are 0.197 and 0.597 respectively. Thus, although these figures imply an apparent misspecification of the standard model without unobserved heterogeneity, this model still gives the same substantive conclusions about the covariate effects on the hazard rates as the model without unobserved heterogeneity.

Table 5 about here

The *comparison area* hazard ratios in columns 2 and 4 of Table 5 show that children are about 84 per cent more likely to die from a communicable disease in the comparison area than in

the ICDDR,B area, a statistically significant difference. The odds of dying from a non-communicable disease are, on the other hand, similar in the two areas; there is only a 10 per cent difference, and this is not significant. These findings are in line with the larger difference between the two areas in mortality from communicable diseases than in mortality from non-communicable diseases, as shown in Figure 1. Boys are significantly more likely than girls in similar families and circumstances to die of either communicable or non-communicable diseases. The difference in the odds ratios between the sexes is larger for communicable diseases than for non-communicable diseases. The figures show that a family's religion plays no significant role in the chances that a child from that family dies from either cause in the first four weeks of life.

Table 5 indicates that if a mother is more highly educated this significantly reduces the hazard of her child dying of either a communicable or a non-communicable disease. The father's education level does not have a clear effect; the only marginally significant coefficient suggests, surprisingly, that, keeping other variables constant, children of fathers with primary education have a higher risk of death from a non-communicable disease than children of fathers with either less education than this, or more. If a father is a day labourer, on the other hand, this substantially and significantly raises the odds of his child dying from a communicable disease.

Consistently with findings reported in the literature (see for example, Bhalotra and van Soest 2008), our figures for *birth order* in Table 5 show that first-born children are at higher risk than higher-order births of neonatal mortality; there are no significant differences amongst birth orders two or higher. The effect of being a first birth is much larger and more significant for non-communicable diseases than for communicable diseases. The hazards tend to fall with the mother's age at the child's birth, although not all the differences are statistically significant. The risk of dying from both communicable and non-communicable diseases in the neonatal period decreases

significantly over time. The risk of dying from a communicable disease fell particularly strongly during the 2000-05 period.

The hazards of dying from both communicable and non-communicable diseases rise significantly as the distance to the nearest health centre increases, and the effect is somewhat larger for communicable than for non-communicable diseases. The other community-specific variable, *no tube-well or piped water*, is not significant.

A monotonically decreasing trend can be observed in the *baseline hazard* of dying from a non-communicable disease: the hazard of dying is greatest on the day of birth but already much lower one day after birth, and continues to decrease for the rest of the neonatal period. In contrast, there is quite a different pattern for communicable diseases: the hazard of dying from one of these diseases increases over the first few days of life and only then starts to fall, gradually. This difference is in line with the results presented in Figure 1 which shows, as we already saw, that deaths attributed to communicable diseases were much less likely to occur in the first few days after birth.

Table 6 presents separate estimates for the ICDDR,B area (columns (1) and (3)) and the comparison area (columns (2) and (4)), using the model with discrete, mother-specific, unobserved heterogeneity. As we saw in Section 5.1, there is no statistical evidence that we should use these separate models rather than the joint model but it may still be instructive to look at some of the estimated hazard ratios when the areas are treated separately. For example, for both communicable and non-communicable diseases the estimated difference in mortality between the sexes is larger in the ICDDR,B area than in the comparison area, suggesting that the additional health services in the ICDDR,B area may have been of particular benefit to male children. The differences are not significant at the 5 per cent level, however. Distance to the nearest health centre seems to have had an impact on mortality only in the comparison area; this is understandable as the distance to the

nearest health centre is typically much greater in the latter area than in the ICDDR,B area, as shown in Table 2. The estimated baseline hazards are similar in the two areas. There seems to have been a rather large difference between the variances of the unobserved heterogeneity terms in both areas, but these are not very precisely determined and, as we saw previously, they had little effect on the other estimates. The estimated correlation coefficients between the two log frailty terms (calculated from the variances and covariances at the bottom of Table 6) are both significantly positive and 0.30 for the ICDDR,B area and 0.67 for the comparison area.

Table 6 about here

5.3. Cumulative incidence functions

The cumulative incidence functions corresponding to the models with deaths attributed to communicable diseases versus non-communicable diseases for a reference births (female child from the 1987-1992 cohort, born into a Muslim family and to a mother aged between 20 and 24, neither parent has any education, father does not work as a day-labourer, water from a tube well or a pipe, living at average distance from the nearest health centre) in both areas are shown in Figure 2. They are substantially different from those for the complete sample in Figure 1 since the socio-economic characteristics of the benchmark case are not representative of the sample average. The lines in the graph confirm our conclusions from the baseline hazards in Table 5: the lines for non-communicable diseases rise much more steeply in the first few days of life than do the lines for communicable diseases. The shape but not the level of the lines for the two causes is similar in both areas. The cumulative number of deaths from both communicable diseases and non-communicable diseases is always higher in the comparison area than in the ICDDR,B area, but the difference

between the two areas is much greater for communicable diseases than for non-communicable diseases.

--- Figure 2 approximately here ---

6. Discussion

We analysed neonatal mortality in Matlab, Bangladesh, over the period 1987–2005, when, as shown in Table 3, the ICDDR,B area, with its high-quality healthcare services, recorded a neonatal mortality rate of 32.3 deaths per 1,000 live births, while the comparison area, a typical rural area in Bangladesh with only standard government health services in place, had a rate of 42.3 neonatal deaths per 1,000 live births. There has been remarkable success in the reduction of neonatal mortality from neonatal tetanus and other immunizable diseases, and this explains the largest part of the reduction in total neonatal mortality in the ICDDR,B area (Bhatia 1989; Baqui et al. 2001). This is reflected in our cause-of-death data, shown in Table 3, which indicates that mortality rates from neonatal tetanus were 0.32 per 1,000 live births in the ICDDR,B area and 2.47 in the comparison area. During the study period, the causes categorized as ‘prematurity/low birth weight’ were the leading causes of neonatal deaths in both areas in our study. This result is consistent with global findings, which point to preterm birth and low birth weight as major causes of neonatal death (Lawn et al. 2006). The age at which a child dies from a disease is an important phenomenon in epidemiological studies. Our findings show that the number of deaths from non-communicable diseases, which occur mainly 1–6 days after birth, fell over the study period, but that the number of deaths attributed to communicable diseases increased. Previous studies of Matlab (Bhatia 1989; Chowdhury et al. 2010) have reported that the ICDDR,B program significantly reduced neonatal mortality in the 24 hours after birth. Our results, however, show that the mortality rate changes with

age in a similar fashion for communicable diseases as well as non-communicable diseases in both areas.

At each day after birth, the cumulative number of deaths from both communicable and non-communicable diseases was larger in the comparison area than in the ICDDR,B area. In both areas male children were more likely than female children to die from both categories of disease. The difference between the sexes was greatest for communicable diseases in the ICDDR,B area, where boys were almost 72 per cent more likely than girls to succumb to these diseases (see Table 6). In the case of deaths from non-communicable diseases, the (rather smaller) male disadvantage was mainly the result of deaths in the 'neonatal conditions' category (NCs in the right hand column of Table 1), that is those caused by birth asphyxia, conditions specific to the neo-natal period, congenital abnormalities, and obstetric complications. These figures are consistent with the findings of other studies which show that infant mortality is inherently higher amongst boys than it is amongst girls, although the situation can be reversed if female children are born into a setting in which they are culturally, socially, or economically disadvantaged (Waldron 1983; Chowdhury et al. 2010). The large disadvantage for boys in the ICDDR,B area suggests that the extensive health services introduced in the ICDDR,B area reduced the influence of any such factors. These findings give additional insight into the causes of differences between the sexes in child mortality, whereas an earlier study could reveal only the overall improvement in the survival of female children survival in the ICDDR,B area in the 1970-1995 period (Datta and Bairagi 2000).

When socioeconomic indicators such as parents' education and father's occupation were kept constant, the risk of death from a non-communicable disease was significantly higher for first-born children and the children of young mothers than for other children, indicating that physiology, rather than socio-economic factors, played a role in these deaths (Bhatia 1989). If a mother was educated this substantially reduced the likelihood that her child would succumb to a neonatal death

from a non-communicable disease, and this is again consistent with general findings in the literature on child mortality (e.g., Dancer et al. 2008).

In both the ICDDR,B area and the comparison area, we found substantial differences between the three cohorts in their risk of dying from both communicable diseases and non-communicable diseases in the neonatal period. In the comparison area the risk of dying from a communicable disease fell particularly strongly during the 2000-05 period, and the number of deaths from non-communicable diseases fell over time from 1987 till 2005 in both areas. We also undertook an analysis using models with four exits: death from a neonatal-related cause (NCs in the right hand column of Table 1), death from low birth weight (LBW), death from a communicable disease (CDs) or death from any another cause ('Other'). This revealed that neonatal-related conditions became the primary leading underlying cause of neonatal death in both of our study areas in the period 2000-05. Chowdhury et al. (2010) also found a relatively high rate of deaths from birth asphyxia in Matlab data from 2003-2004, and suggest that it might be explained by a lack of appropriate resuscitation care. In both areas over the period 2000-05, there was a fall in the risk of dying from low birth weight, communicable diseases, and all other causes, and a remarkably large reduction in deaths in the 'all other causes' category. This may have been due to improvements in verbal autopsy, resulting in more specific causes of death being recorded (see Liu et al. 2011 and Karar et al. 2009).

Our study contributed an important improvement in the models used to study infant mortality by allowing for correlated unobserved heterogeneity, inducing correlation between competing hazards. This was useful because we were unable to include some potentially relevant covariates, such as use of antenatal care, birth practice, and genetic traits in our models, and any of these factors could have led to mother-specific, unobserved variation in the outcomes in which we were interested. In fact, we found clear evidence of unobserved heterogeneity and a positive

correlation between the unobserved heterogeneity terms in the hazards of death from both communicable diseases and non-communicable diseases. Moreover, a model with mother-specific, unobserved heterogeneity performs better than a model with community-specific, unobserved heterogeneity, and a discrete distribution of unobserved heterogeneity fits the data better than a more common, continuous distribution. This suggests that this kind of model is a substantial improvement on the use of single-cause models (see for example Black et al. 2010). Bijwaard and Ridder (2005) demonstrated, using a Monte Carlo study, that, even when unobserved heterogeneity is known to exist, standard estimation methods often do not find evidence of this. A model with significant unobserved heterogeneity is therefore strong evidence that this heterogeneity really plays a role in explaining the neonatal mortality.

We should note, however, that using this advanced procedure, with either mother-specific or community-specific, unobserved heterogeneity, has hardly any effect on the substantive conclusions about how hazard rates vary over the neonatal period or how they relate to any of the covariates. Our conclusions remained unchanged after we disaggregated non-communicable diseases into three groups of cause of death and analysed four competing risks. This may have been because repeated deaths in the same family have become less common in Matlab, which appears to have entered the third stage of the demographic transition (see Karar et al. 2009). Another reason may be that incorporating unobserved heterogeneity is particularly important for longer durations of life and we studied deaths in the neonatal period only.

Our model without any interactions with an area dummy could not be rejected when compared against more general models that allowed for area differences in the covariate effects or baseline hazards. We thus found a higher likelihood of neonatal deaths from both communicable diseases and non-communicable diseases in the comparison area across both socioeconomic groups and time periods. Our findings highlight the contribution to the reduction of neonatal mortality that

can be made by high-quality health services and specific interventions to improve maternal and child health, whether at the antenatal, intra-partum, or postnatal stage. Tetanus toxoid immunization for pregnant women has been particularly important in eliminating deaths from neonatal tetanus, but in order to achieve the reductions in child mortality sought by the fourth Millennium Development Goal, strategies to target both acute respiratory diseases and the causes of neonatal deaths occurring within three days of birth remain necessary.

References

- Adjuik, Martin, Tom Smith, Sam Clark, Jim Todd, Anu Garrib, Yohannes Kinfu, Kathy Kahn, Mitiki Mola, Ali Ashraf, Honorati Masanja, Ubaje Adazu, Jahit Sacarlal, Nurul Alam, Adama Maria, Adjima Gbangou, Eleuther Mwageni, and Fred Binka. 2006. Cause-specific mortality rates in sub-Saharan Africa and Bangladesh, *Bulletin of the World Health Organization* 84(3): 181-188.
- Arulampalam, Wiji, and Sonia Bhalotra. 2006. Sibling death clustering in India: State dependence versus unobserved heterogeneity, *Journal of the Royal Statistical Society A* 169(4): 829-848.
- Baqui, Abdullah H, Ahmed Al Sabir, N. Begum, Shams El Arifeen, S. N. Mitra, and Robert E. Black. 2001. Causes of childhood deaths in Bangladesh: An update, *Acta Paediatrica* 90(6): 682-690.
- Baqui, Abdullah H., G.L. Darmstadt, E.K. Williams, V. Kumar, T.U. Kiran, D. Panwar, V.K. Srivastava, R. Ahuja, R.E. Black and M. Santosham. 2006. Rates, timing and causes of neonatal deaths in rural India: implications for neonatal health programmes, *Bulletin of the World Health Organization* 84(9): 706-713.
- Bhalotra, Sonia, and Arthur van Soest. 2008. Birth spacing, fertility and neonatal mortality in India: dynamics, frailty and fecundity. *Journal of Econometrics*, 143: 274-290.
- Bhatia, Shushum. 1981. Traditional childbirth practices: Implications for a rural MCH program, *Studies in Family Planning* 12 (2): 66-75.
- Bhatia, Shushum. 1989. Patterns and causes of neonatal and postneonatal mortality in rural Bangladesh, *Studies in Family Planning* 20(3): 136-146.
- Bijwaard, Govert E., and Geert Ridder. 2005. Correcting for selective compliance in a reemployment bonus experiment, *Journal of Econometrics* 125(1): 77-111.

- Black, Robert E., Simon Cousens, Hope L. Johnson, Joy E. Lawn, Igor Rudan, Diego G. Bassani, Prabhat Jha, Harry Campbell, Christa Fischer Walker, Richard Cibulskis, Thomas Eisele, Li Liu, and Colin Mathers for the Child Health Epidemiology Reference Group of WHO and UNICEF. 2010. Global, regional, and national causes of child mortality in 2008: A systematic analysis, *The Lancet* 375(9730): 1969-1987.
- Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*, Cambridge UP: New York.
- Chen, Lincoln C., Mizanur Rahman, and A. M. Sarder. 1980. Epidemiology and causes of death among children in a rural area of Bangladesh, *International Journal of Epidemiology* 9(1): 25-33.
- Chowdhury, Hafizur Rahman, Sandra Thompson, Mohammad Ali, Nurul Alam, Md. Yunus, and Peter Kim Streatfield. 2010. Causes of neonatal deaths in a rural subdistrict of Bangladesh: Implications for intervention, *Journal of Health Population Nutrition* 28(4): 375-382.
- Conde-Agudelo, Agustin, Jose M Belizan, and Jose Diaz-Rossello. 2003. Kangaroo mother care to reduce morbidity and mortality in low birth weight infants (Cochrane Review), in: *The Cochrane Library, Issue 4*, Oxford: Update Software.
- Cornfield, Jerome. 1957. The estimation of the probability of developing a disease in the presence of competing risks, *Disease Probability* 47(5): 601-607.
- Coviello, Vincenzo, and May Boggess. 2004. Cumulative incidence estimation in the presence of competing risks, *The Stata Journal* 4(2): 103-112.
- Cox, David R. 1959. The analysis of exponentially distributed life-times with two types of failures, *Journal of the Royal Statistical Society B* 21(2): 411-421.
- Dancer, Diane, Anu Rammohan and Murray D. Smith. 2008. Infant mortality and child nutrition in Bangladesh. *Health Economics* 17(9): 1015-1035.

- DasGupta, Monica. 1990. Death clustering, mothers' education and the determinants of child mortality in rural Punjab, India, *Population Studies* 44(3): 489-505.
- Datta, Ashish Kumar, and Radheshyam Bairagi. 2000. Improvement in female survival: A quiet revolution in Bangladesh, *Asia Pacific Population Journal* 15(1): 19-40.
- DaVanzo, Julie, Lauren Hale, Abdur Razzaque, and Mizanur Rahman, 2008. The effects of pregnancy spacing on infant and child mortality in Matlab, Bangladesh: How they vary by the type of pregnancy outcome that began the interval, *Population Studies* 62(2) 131-154.
- Fauveau, Vincent, Michael A. Koenig, and Bogdan Wojtyniak. 1991. Excess female deaths among rural Bangladeshi children: An examination of cause-specific mortality and morbidity, *International Journal of Epidemiology* 20(3): 729-735.
- Fauveau, Vincent, Bogdan Wojtyniak, Hafizur Rahman Chowdhury, and Abdul Majid Sarder. 1994. Assessment of cause of death in the Matlab Demographic Surveillance System, in: Fauveau, V. (ed.), *Matlab: Women, Children and Health*. Dhaka: International Centre for Diarrhoeal Diseases Research, pp. 65-86.
- Fine, Jason P., and Robert J. Gray. 1999. A proportional hazards model for the subdistribution of a competing risk, *Journal of the American Statistical Association* 94(446): 496-509.
- Goldberg, I.D., Morton L. Levin, Paul R. Gerhardt, Vincent H. Handy, and Rita E. Cashman. 1956. The probability of developing cancer, *Journal of the National Cancer Institute* 17(2): 155-173.
- Hale, Lauren, Julie DaVanzo, Abdur Razzaque, and Mizanur Rahman. 2006. Why are infant and child mortality rates lower in the MCH-FP Area of Matlab, Bangladesh? Are the differences due to differences in reproductive patterns?, *Studies in Family Planning* 37(4): 281-292.
- Heckman, James J., Joseph V. Hotz, and James R. Walker. 1985. New evidence on the timing and spacing of births, *American Economic Review* 75(2): 179-184.

- ICDDR,B. 2006. *Health and Demographic Surveillance System-Matlab: Registration of health and demographic events 2004*, Scientific Report 93. Dhaka: Centre for Health and Population Research.
- Jones, Gareth, Richard W Steketee, Robert E. Black, Zulfiqar A. Bhutta, Saul S. Morris, and the Bellagio Child Survival Study Group. 2003. How many child deaths can we prevent this year?, *The Lancet* 362(9377): 65-71.
- Karar, Zuniad A., Nurul Alam, and Peter Kim Streatfield. 2009. Epidemiological transition in rural Bangladesh, 1986-2006, *Global Health Action* 2(1): 1-9.
- Lawn, Joy E, Simon Cousens, and Jelka Zupan. 2005. 4 million neonatal deaths: When? Where? Why?, *The Lancet* 365(9462): 891-900.
- Lawn, Joy E, Katarzyna Wilczynska-Ketende, and Simon Cousens. 2006. Estimating the causes of 4 million neonatal deaths in the year 2000, *International Journal of Epidemiology* 35(3): 706-718.
- Liu, Li, Qingfeng Li, Rose A. Lee, Ingrid K. Friberg, Jamie Perin, Neff Walker and Robert E. Black. 2011. Trends in causes of death among children under 5 in Bangladesh, 1993-2004: An exercise applying a standardized computer algorithm to assign causes of death using verbal autopsy data, *Population Health Metrics* 9(1): 1-11.
- Manton, Kenneth G., Eric Stallard, and James W. Vaupel. 1981. Methods for the mortality experience of heterogeneous populations, *Demography* 18(3): 389-410.
- Mosley, W. Henry, and Lincoln C. Chen. 1984. An analytical framework for the study of child survival in developing countries, *Population Development Review* 10(Suppl.): 25-45.
- National Institute of Population Research and Training (NIPORT), Mitra and Associates, and Orc Macro. 2005. *Bangladesh Demographic Health Survey 2004*. Dhaka, Bangladesh and Calverton, Maryland: National Institute of Population Research and Training.

- Phillips, James F., Ruth Simmons, J. Chakraborty, and A. I. Chowdhury. 1984. Integrating health services into an MCH-FP program: Lessons from Matlab, Bangladesh, *Studies in Family Planning* 15(4): 153-161.
- Phillips, James F., Ruth Simons, Michael A. Koenig, and J. Chakraborty. 1988. Determinants of reproductive change in a traditional society: Evidence from Matlab, Bangladesh. *Studies in Family Planning* 19(6): 313-334.
- Ronsmans, Carine, Mahbub E. Chowdhury, Shushil Kanta DasGupta, Anisuddin Ahmed, and Marge Koblinsky. 2010. Effect of parent's death on child survival in rural Bangladesh: a cohort study, *The Lancet* 375 (9730): 2024-2030.
- Saha, Unnati R., and Arthur van Soest (2011), Infant mortality in rural Bangladesh: State dependence vs. unobserved heterogeneity, *Population Studies* 65(3): 273-287.
- Sastry, Narayan. 1997. Family-level clustering of childhood mortality risk in Northeast Brazil, *Population Studies* 51(3): 245-261.
- Self, Steven G., and Kung-Yee Liang. 1987. Asymptotic properties of maximum-likelihood estimators and likelihood ratio tests under nonstandard conditions, *Journal of the American Statistical Association* 82(398): 606-610.
- United Nations. 2001. Road map towards the implementation of the United Millennium Declaration: *Report of the Secretary-General, General Assembly, 56th session*. New York: United Nations.
- Waldron, Ingrid. 1983. Sex differences in human mortality, *Social Science and Medicine* 17(6): 321-333.
- World Health Organization. 1977. *Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death. 9th revision*. Geneva: World Health Organization.

Table 1 Cause-of-death categories used in the models and their origin in assignment of causes by the Health and Demographic Surveillance System, Matlab, Bangladesh, 1987-2005

Cause-of-death codes (ICD9), [ICD10]	Cause-of-death label	Cause-of-death categories used in Table 3	Categories used in two- exit model (Table 4 & 5)	Categories used in four- exit model (results not shown in text)
(190, 192, 193,452,458), [P05,P07, P08]	Preterm delivery/low birth weight	Low birth weight	Non-communicable diseases	Low birth weight
(454),[P21]	Birth asphyxia	Birth asphyxia		Non-communicable diseases
(457), [P22-P29, P35,P36,P51, P59, P76,P80]	Infections, respiratory and cardiovascular disorder specific to the perinatal period, Digestive system disorders of foetus, Haemorrhagic.	Neonatal related conditions		
(449), [Q01, Q02, Q03,Q24, Q35,Q37, Q42,Q45,Q89]	Congenital abnormalities	Congenital abnormalities		
(453), [P15]	Birth trauma/cord haemorrhage	Birth trauma/cord haemorrhage		
(451),[P00,P01,P02,P03]	Obstetric complications or foetus and newborn affected by maternal factors and by complications of pregnancy, labour and delivery	Obstetric complications of new born (OBSCOMP)		
[P90-P96]	Other disorders originated in the perinatal period	unspecified		Other
(990,998,999),[R34,R96,R99]	Unspecified causes			
(450,459), [R95],	Sudden death			
(293,344,420,460,461,552,55, 5,559,691,738) [K75, W75,X91, S27]	Other specific	Other specific		
(010,013,014)	Acute watery diarrhoea, dysentery, acute non-watery diarrhoea	Diarrhoea, dysentery	Communicable diseases	Communicable diseases
(038, 046), [G03]	Septicaemia, viral hepatitis	Septicaemia, hepatitis, meningitis		
(321, 325,328,191),[J11, J18]	Pneumonia, Acute lower respiratory infections, Pneumonia with diarrhoea, Pneumonia severe, influenza and pneumonia	Acute respiratory infections		
[A41,B01, A03]	Other bacterial diseases, viral infections characterized by skin and mucous membrane lesions, intestinal infectious diseases	Other infectious diseases		
(456)	Immunizable diseases: neonatal tetanus	Extended Program on Immunization		

Note: The coding schemes shown above follow the verbal autopsy codings applied to neonatal deaths in the Matlab area by Bhatia (1989), Chowdhury et al. (2010) and the Child Health Epidemiology Reference Group of WHO and UNICEF (see, for example, Liu et al. 2011) .

Source: Health and Demographic Surveillance System, Matlab, Bangladesh

Table 2 Summary statistics of explanatory variables^{1,2}; ICDDR,B area and comparison area

Variable	ICDDR,B area Per cent of all births (N=49,537)	Comparison area Per cent of all births (N=57,830)
Sex of index child		
Male	50.78	50.76
Child's birth cohort		
Before 1993	32.86	36.53
1993-1999	34.63	34.26
2000-2005	32.51	29.22
Birth order		
1	30.97	26.65
2-3	43.07	38.23
4 +	25.96	35.12
Religion: Non-Muslims	14.67	8.91
Mother's education		
No education	49.00	53.10
Primary education	25.15	25.05
At least secondary education	25.86	21.85
Mother's age at birth		
<20	12.19	11.54
20-24	33.31	33.13
25-29	28.55	28.39
30 +	25.94	17.8
Father's education		
No education	57.26	58.72
Primary education	21.70	23.00
At least secondary education	21.04	18.28
Father's occupation		
Day labourer	15.75	18.37
Source of drinking water		
No tube-well or piped water	25.31	24.10
Distance to health centre in km	1.86 (0.96) ^b	7.29 (3.96)

Note: ¹ Percentages of outcome 1 for all dummy variables;

² Mean and standard deviation (in parentheses) as this is a continuous variable.

Source: As for Table 1

Table 3 Distribution of cause-specific neonatal deaths, ICDDR,B area and comparison area, Matlab, Bangladesh, 1987-2005

Cause of death	ICDDR,B area (N=49,537)		Comparison area (N=57,830)	
	Number of deaths	Deaths per 1,000 live births	Number of deaths	Deaths per 1,000 live births
Communicable diseases	210	4.23	536	9.27
Septicaemia/Meningitis/Hepatitis	7	0.14	1	0.02
Acute respiratory infections	173	3.49	376	6.50
Diarrheal diseases	14	0.28	16	0.28
Neonatal tetanus and other immunizable diseases	16	0.32	143	2.47
Non-communicable diseases (Non-communicable diseases)	1,391	28.09	1,910	33.04
Congenital abnormality	39	0.79	42	0.73
Prematurity/low birth weight	737	14.88	894	15.46
Birth asphyxia	84	1.7	90	1.56
Obstetric complications of new born	140	2.83	225	3.89
Birth trauma/cord haemorrhage	24	0.48	25	0.43
Other neonatal related conditions ¹	74	1.49	97	1.68
Miscellaneous ²	27	0.55	27	0.47
Diagnosis not possible ³	266	5.37	510	8.82
Total	1,601	32.31	2,446	42.31

Notes: See also Table 1 for details on the coding of causes of death

¹ Includes neonatal infections, respiratory and cardiovascular specific disorders to the perinatal period

² Includes fever, jaundice, intestinal obstruction, Oedemas, external cause (injury), homicide

³ Includes sudden infant death, unspecified cause, other disorders in the perinatal period

Source: As for Table 1.

Table 4 Goodness of fit of models used to account for differences between ICDDR, B area and comparison area in neonatal mortality, Matlab, Bangladesh, 1987-2005

	1.		2.		3.	
	Joint model with area dummy only		Areaspecific β_j but joint same baseline hazards and unobserved heterogeneity parameters		All parameters area-specific (separate models for the two areas)	
Model specification	Log likelihood	N of parameters.	Log likelihood	N of parameters.	Log likelihood	N of parameters.
1. No unobserved heterogeneity	-29,127.28	44	-29,107.25	76	-29,102.11	84
2. Discrete mother-specific effects	-29,013.27	49	-28,993.70	81	-28,984.80	94
<i>Benchmark specification</i>						
3. Discrete village-specific effects ¹	-29,120.38	49	-29,100.92	81	²	
4. Log normally distributed mother-specific effects	-29,029.99	47	-29,010.27	79	-29,002.25	90

Notes:

¹ Discrete random effects, but village-specific instead of mother-specific effects

² The maximum likelihood algorithm did not converge

Source: As for Table 1

Table 5 Estimated hazard ratios of competing-risks model for neonatal deaths from communicable versus non-communicable diseases, joint analysis, Matlab, Bangladesh, 1987-2005

Variable	Communicable diseases				Non-communicable diseases			
	Without unobserved Heterogeneity		With unobserved heterogeneity		Without unobserved Heterogeneity		With unobserved heterogeneity	
	(1)		(2)		(3)		(4)	
Male	1.33**	(0.10)	1.34**	(0.10)	1.17**	(0.04)	1.18**	(0.04)
Hindu	1.16	(0.13)	1.16	(0.13)	1.06	(0.05)	1.06	(0.06)
Mother's education level								
At least primary	0.81*	(0.08)	0.81*	(0.08)	0.89*	(0.04)	0.88*	(0.04)
At least secondary	0.80	(0.10)	0.78*	(0.10)	0.70**	(0.04)	0.66**	(0.04)
Mother's age at birth								
<20 years	1.19	(0.15)	1.18	(0.10)	1.27**	(0.06)	1.27**	(0.07)
25-29 years	0.72**	(0.08)	0.74*	(0.08)	0.95	(0.05)	0.98	(0.05)
≥30	0.78	(0.10)	0.82	(0.11)	1.09	(0.07)	1.16*	(0.08)
Birth order								
2-3	0.78*	(0.08)	0.75*	(0.08)	0.52**	(0.03)	0.49**	(0.02)
≥4	0.93	(0.13)	0.84	(0.12)	0.57**	(0.04)	0.49**	(0.03)
Father's education level								
At least primary	0.96	(0.09)	0.97	(0.09)	1.10*	(0.05)	1.12*	(0.05)
At least secondary	0.79	(0.09)	0.80	(0.10)	0.95	(0.05)	0.96	(0.05)
Father day labourer	1.49**	(0.13)	1.50**	(0.13)	1.07	(0.05)	1.06	(0.05)
No tube-well or piped water	1.15	(0.10)	1.16	(0.11)	1.08	(0.04)	1.08	(0.05)
Distance to health centre (kilometres)¹	1.029*	(0.01)	1.029*	(0.01)	1.016*	(0.01)	1.017*	(0.01)
Birth cohort child								
1993-1999	0.91	(0.08)	0.90	(0.09)	0.84**	(0.04)	0.83**	(0.04)
2000-2005	0.66**	(0.07)	0.65**	(0.07)	0.75**	(0.04)	0.74**	(0.04)
Comparison area	1.82**	(0.19)	1.84**	(0.20)	1.08	(0.05)	1.10	(0.06)
Baseline intensity								
Day 0	0.60	(0.17)	0.57*	(0.16)	35.67**	(1.73)	33.76**	(1.67)
Day 1	1.82**	(0.31)	1.76**	(0.30)	16.54**	(0.95)	15.96**	(0.93)
Days 2	2.13**	(0.33)	2.08**	(0.32)	11.86**	(0.76)	11.56**	(0.74)
Days 3-6	1.71**	(0.16)	1.68**	(0.16)	4.14**	(0.24)	4.08**	(0.24)
Constant	-8.75**	(0.15)	-7.82**	(0.29)	-7.75**	(0.07)	-6.83**	(0.40)
Unobserved heterogeneity								
Variance	-	-	0.20	(0.07)	-	-	0.60	(3.01)
Covariance	-	-	0.16	(0.06)	-	-	-	-

Notes: The hazard ratios presented are equal to $\exp(\beta_j)$ in equation (1).

The reference individual is a female child from the 1987-1992 cohort, born into a Muslim family and to a mother aged between 20 and 24, and living in the ICDD/RB area. Neither of the child's parents has any education but her father does not work as a daylabourer. The family get their water from a tube well or pipe, and live an average distance from the nearest health centre., baseline intensity 7-28 days.

¹ The distance to the nearest health centre is centred around its mean value in the ICDDR,B area and the comparison area; * $2 < t\text{-value} < 3$ (0.0027 < p-value < 0.045), ** $t\text{-value} > 3$ (p-value < 0.0027); standard errors are in parentheses

Source: As for Table 1

Table 6 Estimated hazard ratios of competing-risks model for neonatal deaths from communicable and non-communicable diseases, separate analyses for the ICDDR,B area and comparison area, Matlab, Bangladesh, 1987-2005

Variable	Communicable diseases				Non-communicable diseases			
	ICDDR,B area		Comparison area		ICDDR,B area		Comparison area	
Male	1.72**	(0.27)	1.22*	(0.11)	1.26**	(0.07)	1.12*	(0.05)
Hindu	0.97	(0.21)	1.32	(0.19)	1.00	(0.08)	1.15	(0.09)
Mother's education level								
At least primary	0.83	(0.17)	0.80	(0.10)	0.86*	(0.06)	0.90	(0.06)
At least secondary	0.79	(0.19)	0.77	(0.12)	0.67**	(0.06)	0.69**	(0.06)
Mother's age at birth								
<20 years	1.34	(0.32)	1.13	(0.17)	1.22*	(0.10)	1.31**	(0.09)
25-29 years	0.63*	(0.13)	0.79	(0.11)	0.96	(0.08)	1.00	(0.07)
≥30	0.59	(0.17)	0.91	(0.14)	1.23	(0.13)	1.12	(0.10)
Birth order								
2-3	0.91	(0.19)	0.68*	(0.09)	0.48**	(0.04)	0.50**	(0.03)
≥4	1.13	(0.36)	0.74	(0.13)	0.51**	(0.06)	0.49**	(0.04)
Father's education level								
At least primary	0.85	(0.18)	1.02	(0.11)	1.20*	(0.09)	1.06	(0.06)
At least secondary	0.74	(0.18)	0.83	(0.12)	0.93	(0.08)	0.97	(0.07)
Father day labourer	1.20	(0.23)	1.62**	(0.17)	0.97	(0.08)	1.14*	(0.07)
No tube-well or piped water	1.36	(0.24)	1.03	(0.12)	1.12	(0.08)	1.05	(0.07)
Distance to health centre (km)	1.01	(0.08)	1.03*	(0.01)	1.00	(0.03)	1.02*	(0.01)
Birth cohort child								
1993-1999	1.05	(0.19)	0.81	(0.09)	0.73**	(0.05)	0.90	(0.05)
2000-2005	0.72	(0.15)	0.61**	(0.09)	0.72**	(0.05)	0.75**	(0.05)
Baseline intensity								
Day 0	0.63	(0.35)	0.75*	(0.25)	36.97**	(2.89)	31.82**	(2.04)
Day 1	1.45	(0.53)	1.86**	(0.37)	17.99**	(1.61)	14.73**	(1.14)
Days 2	3.42**	(0.85)	1.60**	(0.32)	11.82**	(1.21)	11.36**	(0.94)
Days 3-6	2.03**	(0.36)	1.55**	(0.17)	4.26**	(0.39)	3.94**	(0.30)
Constant	-7.98**	(0.48)	-8.01**	(0.54)	-7.34**	(1.87)	-6.72**	(0.28)
Unobserved heterogeneity								
Variance	0.16	-	0.37	-	1.42	-	0.39	-
Covariance	0.15	-	0.25	-				

Notes: As for Table 5

Source: As for Table 1

Figure 1: Non-parametric estimates of cause-specific neonatal deaths, ICDDR,B area and comparison area, Matlab, Bangladesh, 1987-2005

Source: As for Table 1

Figure 2: Model-based estimates of neonatal deaths from communicable and non-communicable diseases for benchmark cases¹, ICDDR,B area and comparison area, Matlab, Bangladesh, 1987-2005

Source: As for Table 1

¹ Footnote: The reference individual is a female child from the 1987-1992 cohort, born into a Muslim family and to a mother aged between 20 and 24, and living in the ICDDR,B area. Neither of the child's parents has any education but her father does not work as a day-labourer. The family get their water from a tube well or a pipe, and live an average distance from the nearest health centre