Estimating the causes of 4 million neonatal deaths in the year 2000

Joy E Lawn,1,2 Katarzyna Wilczynska-Ketende2,3 and Simon N Cousens4*

Accepted 22 February 2006

Background Information on cause-of-death is lacking for 98% of the world’s 4 million neonatal deaths that occur in countries with inadequate vital registration (VR). Our aim was to estimate, by country for the year 2000, the distribution of neonatal deaths across programme-relevant causes including: asphyxia, preterm birth, congenital abnormalities, sepsis/pneumonia, neonatal tetanus, diarrhoea, and ‘other’.

Methods Two sources of neonatal cause-of-death data were examined: VR datasets for countries with high coverage (>90%), and published and unpublished studies identified through systematic searches. Multinomial regression was used to model the distribution of neonatal deaths. A VR-based model was used to estimate the distribution of causes of death for 37 low-mortality countries without national data. A study-based model was applied to obtain estimates for 111 high-mortality countries. Uncertainty estimates were derived using the jackknife approach.

Results Data from 44 countries with VR (96,797 neonatal deaths) and from 56 studies (29 countries, 13,685 neonatal deaths) met inclusion criteria. The distribution of reported causes of death varied substantially between countries and across studies. Based on 193 countries, the major causes of neonatal death globally were estimated to be infections (sepsis/pneumonia, tetanus, and diarrhoea, 35%), preterm birth (28%), and asphyxia (23%). Regional variation is important. Substantial uncertainty surrounds these estimates.

Conclusions This exercise highlights the lack of reliable cause-of-death data in the settings in which most neonatal deaths occur. Complex statistical models are not a panacea. Representative data with comparable case definitions and consistent hierarchical cause-of-death attribution are required.

Keywords Infant, newborn, mortality, cause of death

It is estimated that each year 4 million children die in the first 4 weeks of life—the neonatal period—a global average of 30 neonatal deaths per 1000 livebirths.1 The fourth Millennium Development Goal (MDG-4) aspires to reduce under-5 child mortality to close to 30 child deaths per 1000 livebirths by the year 2015. Without substantial reductions in the global neonatal mortality rate (NMR) MDG-4 will not be achieved.2

1 Saving Newborn Lives/Save the Children-USA, 11 South Way, Pinelands, Cape Town 7405, South Africa.
2 Honorary Research Fellow, Centre for International Child Health, Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK.
3 Interagency Group on Breastfeeding Monitoring, UNICEF, UK.
4 Infectious Diseases Epidemiology Unit, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK.
5 Corresponding author. E-mail: Simon.Cousens@lshtm.ac.uk
6 On behalf of the CHERG neonatal group. The study was designed by Joy Lawn and Simon Cousens. Joy Lawn designed the VR analysis, performed the searches, screened and abstracted data, collaborated in the modelling strategy, and wrote the manuscript. Simon Cousens designed and carried out the modelling and contributed substantially to the manuscript. Kate Wilczynska assisted with the searches, abstracted data, particularly covariates, supported the correspondence with study authors, and reviewed the manuscript. The Child Health Epidemiology Reference Group (CHERG) neonatal group was involved in a series of study design and evaluation meetings and reviewed the manuscript. Members include Zulfiqar Bhutta (Aga Khan University, Karachi), Robert Black (CHERG chair person, Bloomberg School of Public Health, Johns Hopkins, Baltimore, MD), Karen Edmond (London School of Hygiene and Tropical Medicine), Jose Martines, Kenji Shibuya, Martin Weber, and Jelka Zupan (WHO, Geneva).
Many neonatal deaths are preventable with existing low-cost interventions, but to make the best use of limited resources, planners and policy makers require reliable cause-of-death information. However, 99% of the world’s neonatal deaths occur in low-income and middle-income countries, few of which have high vital registration (VR) coverage. The only option currently available to meet this gap in information regarding the vast majority of neonatal deaths is estimation. Estimates are available for selected single causes of neonatal deaths such as those due to birth asphyxia or intrapartum events and neonatal tetanus. Before 2005, the World Health Organization (WHO), provided little detail with respect to the causes of neonatal deaths, with 2.6 million neonatal deaths grouped together as ‘perinatal causes’. This was the biggest single category of deaths in the global burden of disease tables and included several distinct causes of death with differing programmatic solutions. Neonatal infections, the single largest cause of neonatal deaths globally, were not included in the perinatal causes group and were not distinguishable from infections after the neonatal period, despite the need for alternative prevention and treatment strategies. Furthermore, the data inputs and methods for these estimates were not available.

The science of systematic reviews of interventions is advanced, with guidelines for search strategies and inclusion criteria, for example, in Cochrane reviews (http://www.cochrane.org/resources/handbook/). The science of disease burden is less advanced and at times controversial. Comprehensive searches, descriptions of modelling, and estimation of uncertainty are becoming the norm. New estimation approaches that constrain all the major causes of death in a given age band to fit the total number of deaths in that given age band are more attractive than attempting to combine multiple single cause estimates generated through varying methods. However the methodological and statistical challenges are considerable, and there may be other disadvantages.

Our aim was to provide for 193 countries in the year 2000, systematic estimates, with associated uncertainty, of the distribution of neonatal deaths for programme-relevant causes; birth asphyxia, preterm birth, congenital abnormalities, sepsis/pneumonia, neonatal tetanus, and diarrhoea, with a residual category of ‘other’ including specific but less common causes of neonatal death.

Materials and methods

Overview

Two sources of cause-of-neonatal-death data were used: VR data, and published and unpublished reports of research studies (study data). The search strategy and inclusion criteria are described in Table 1. National level estimates of proportionate causes were obtained in one of three ways. For 45 countries with high coverage (>90%) VR data, these data were used. Second, for 37 countries with low NMRs but without high coverage VR systems, estimates were obtained by applying a multinomial regression model developed using high coverage VR data. Finally for 111 countries with higher NMRs and without VR data, estimates were obtained by applying a multinomial regression model developed using the study data. For both the VR-data-based and the study-data-based models, national predictions were derived by applying the models to national level covariate data for the year 2000.

For the purposes of our analysis we identified six cause-of-death categories, plus one residual category, based on the following considerations: expected public health importance, differing implications for intervention, and the ability to distinguish between them in low resource settings. Thus, since pneumonia in a neonate cannot be distinguished on clinical examination from septicaemia or meningitis, and because case management is similar for all three conditions, one category, subsequently referred to as ‘sepsis/pneumonia’, was used for all three causes. The category ‘preterm’ included only deaths directedly attributed to prematurity and to specific complications of preterm birth such as surfactant deficiency, but not all deaths in preterm infants. The cause-of-death categories and case definitions used are summarized in Table 2.

Input data

VR data

The WHO supplied a database of VR data, covering 83 countries with two different International Classification of Disease coding systems (ICD9 and ICD10). We used the data from the year closest to the year 2000. If the annual number of neonatal deaths in the country was <500, we used the average for the 3 years closest to the year 2000. Excel spreadsheets (Microsoft XP, 2000) and Stata version 8 programs (Stata Corporation, College Station, TX, USA) were written to link the 20 000 possible codes in ICD10 and 10 000+ codes in ICD9 with the seven cause-of-death categories selected. An ICD9 to ICD10 translation guide was used to maximize consistency between the two classification systems.

Study data

We performed systematic searches of the published literature and made extensive attempts to identify non-English language publications (Table 1) and unpublished datasets. After applying inclusion criteria (Table 1), data on numbers of neonatal deaths by cause and on potential explanatory variables were abstracted by two independent abstractors using a standard form. Deaths were allocated among our seven cause-of-death categories using the authors’ cause-of-death attribution. If authors gave more than one cause of death per neonate then a fixed hierarchy was applied, following ICD rules where possible (Table 2). For example, a death in a neonate with a neural tube defect and infection was classified as due to congenital abnormality. We contacted the authors for additional data regarding missing or unclear causes. For example, if a neonatal death was attributed to ‘feeding difficulties’ the authors were asked to supply additional information regarding the death to allow allocation to a standard category. Deaths from unknown causes were excluded from subsequent analysis, but if more than 25% of deaths were unknown the study was excluded (Table 1).

We also abstracted data for a range of variables that might explain the proportional distribution of causes in a study (Table 3). These variables related to the study site and study design/methods, to the overall NMR (e.g. low birth weight rate,
skilled-attendant coverage), or to specific causes of neonatal deaths [e.g. tetanus-toxoid coverage (TT2)]. One limitation on the variables was the requirement that national covariate data would be available for all countries for prediction purposes. Some covariates of interest, such as coverage of emergency obstetric care or early post-natal/newborn care, are not routinely collected. We considered it important to identify values for explanatory variables, which applied to the study population, or as close to it as possible, in view of the possibility that study populations might not be representative of national populations. For example, in three studies from The Gambia, tetanus accounts for 1% of neonatal deaths. Locally high TT2 coverage is important in explaining this.

We wrote to 55 authors to obtain additional information on causes of death and local explanatory variables. When study population-specific explanatory data were not available we used other sources such as Demographic and Health Surveys (DHS, www.measuredhs.com) and local programme reports. We identified local or regional data for 90% of the 56 studies included for all the indicators except TT2+ (83%) and those which are by definition national such as gross domestic product (GDP) per capita.

**Modelling**

Modelling was performed separately for the two datasets (VR and study data). All analyses were performed using Stata version 8 software. A two-step approach was applied to each dataset.

**Step one:** One cause was identified as a ‘baseline’ cause for each dataset, and the logarithm of the ratio of each of the other causes to the baseline cause was calculated, adapting the method applied by Morris *et al*. Ordinary logistic regression was used to develop models for each ratio. For the VR data, with only relatively small variations in NMR, we used a forward stepwise approach based on statistical significance testing, at the 5% level. For the study data, models based on statistical significance alone resulted in multiple parameters and, therefore, we included only variables that we expected a priori to be associated with each ratio (Table 3) and for which the parameter estimate had the expected sign and explained some variability. For example, we expected that the tetanus:asphyxia ratio would be associated with the coverage of tetanus-toxoid immunization, with the ratio decreasing as coverage increases.

**Step two:** The explanatory variables identified using the log ratio models as described above were fitted simultaneously in a multinomial model including all causes to obtain parameter estimates for use in predictions. To allow for within-data source correlations, robust rather than model-based standard errors were used and studies were given a weight proportional to the square root of the number of deaths on which they had data. This weighting is intermediate between giving equal weight to each study or equal weight to each death.
For countries with high VR coverage (>90%), we used the reported distribution of causes of death (45 countries, 2.4% of neonatal deaths). The VR model was used to predict the proportional distribution of causes of death in countries without high coverage VR but with an NMR of <10 per 1000 (all regions) or with an NMR of <15 per 1000 for countries in the European (EURO) and American (AMRO) regions as defined by WHO (37 countries, 2.4% of neonatal deaths). EURO and AMRO regions had VR data points in the NMR range 10–15 per 1000. For all other countries (111 countries, 95.2% of neonatal deaths), predictions were derived using the study data model. For both models, prediction of the distribution of causes of neonatal death at national level required national level covariate data. We used data for the year 2000 from global databases of UNICEF, WHO, and the World Bank (Table 3). We then applied the predicted proportions to WHO estimates of the total number of neonatal deaths occurring in each country to obtain estimates of the number of deaths by cause for each country.

Uncertainty
Uncertainty estimates were obtained using the jackknife approach. Each study or country was removed in turn from the multinomial model estimation step and the predictions for that study/country obtained using the remainder of the data. The distribution of the differences between the observed and estimated log ratios obtained provides an estimate of the standard error of out-of-sample predictions. We used Monte Carlo simulation (10,000 simulations) to randomly perturb country-level estimates based on these standard errors and took the 2.5th and 97.5th centiles to provide an indication of the level of uncertainty in our estimates. This does not capture all the potential sources of variability and uncertainty, such as uncertainty around the number of neonatal deaths in a country.

Results
VR data inputs
A total of 45 countries met the initial inclusion criteria (Figure 1). We excluded Mauritius from the estimation process.
as it was the only African country with high coverage VR data and we did not consider that Mauritius was representative of other African countries given the NMR is 12 per 1000 livebirths. Thus the VR model was based on data on 96 797 deaths from 44 countries, which together account for ~2% of the estimated global total of neonatal deaths (Supplementary Table 2). NMRs ranged from 2 to 18 per 1000 livebirths. There were no reported neonatal tetanus deaths in these countries and very few neonatal deaths due to diarrhoea (290 or 0.3%) so we restricted our modelling to the remaining five causes of death (preterm, sepsis/pneumonia, asphyxia, congenital, and ‘other’) with the small number of diarrhoea deaths allocated to the sepsis/pneumonia (infection) category. The annual number of deaths per country ranged from 12 (Iceland) to 23 603 (Mexico). The recorded distribution of the different causes of death varied between countries particularly for preterm and congenital (Figure 2a). The most common cause of death (preterm) was chosen as the baseline cause for modelling.

### Study data inputs

After applying inclusion criteria, we identified 48 studies and 8 unpublished databases reporting a total of 13 685 deaths with

<table>
<thead>
<tr>
<th>Model</th>
<th>Cause ratio</th>
<th>Variables</th>
<th>Source, all around the year 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vital registration model</td>
<td>All ratios</td>
<td>WHO subregion</td>
<td>WHO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Neonatal and Infant Mortality Rate</td>
<td>UNICEF/WHO estimates</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GDP/GNI per capita</td>
<td>World Bank</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANC coverage</td>
<td>DHS and UNICEF</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Proportion of deliveries attended by a skilled attendant</td>
<td>DHS and UNICEF</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low birth weight rate</td>
<td>UNICEF/WHO estimates</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female literacy rate</td>
<td>UNICEF</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Child survival index</td>
<td>WHO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total fertility rate</td>
<td>United Nations Population Division</td>
</tr>
<tr>
<td>Study data model</td>
<td>All ratios</td>
<td>Study covered first 7 or first 28 days of life</td>
<td>Higher proportion of asphyxia in first 7 days</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Year of study</td>
<td>Evidence of trend</td>
</tr>
<tr>
<td>Preterm:asphyxia</td>
<td>Low birth weight rate</td>
<td>Preterm increasing with LBW rate</td>
<td>Preterm increasing as skilled attendance increases</td>
</tr>
<tr>
<td></td>
<td>Institutional delivery/skilled attendance rate</td>
<td>Asphyxia decreasing as skilled attendance increases</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Study assessed gestational age</td>
<td>Preterm increased if study recorded gestational age</td>
<td></td>
</tr>
<tr>
<td>Sepsis/pneumonia:asphyxia</td>
<td>NMR</td>
<td>Sepsis/pneumonia increasing as NMR increases</td>
<td>Sepsis/pneumonia increasing as LBW rate increases</td>
</tr>
<tr>
<td></td>
<td>Low birth weight rate</td>
<td>Sepsis/pneumonia increasing as LBW rate increases</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BCG coverage shortly after birth</td>
<td>Sepsis/pneumonia decreases as BCG coverage increases (BCG coverage being an indicator of early postnatal care contact)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female literacy rate</td>
<td>Sepsis/pneumonia decreases as literacy increases</td>
<td></td>
</tr>
<tr>
<td>Congenital:asphyxia</td>
<td>NMR</td>
<td>Asphyxia increases as NMR increases</td>
<td>Asphyxia increases as skilled attendance decreases</td>
</tr>
<tr>
<td></td>
<td>Institutional delivery/skilled attendance rate</td>
<td>Asphyxia increases as skilled attendance decreases</td>
<td></td>
</tr>
<tr>
<td></td>
<td>WHO subregion</td>
<td>Congenital higher in EMRO</td>
<td></td>
</tr>
<tr>
<td>Diarrhoea:asphyxia</td>
<td>NMR</td>
<td>Diarrhoea increases with increasing NMR</td>
<td>Diarrhoea increases with increasing LBW rate</td>
</tr>
<tr>
<td></td>
<td>Low birth weight rate</td>
<td>Diarrhoea increases with increasing LBW rate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BCG coverage shortly after birth</td>
<td>Diarrhoea decreases as BCG coverage increases</td>
<td></td>
</tr>
<tr>
<td>Tetanus:asphyxia</td>
<td>TT2+ coverage</td>
<td>Tetanus decreases as TT2+ coverage increases</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NMR</td>
<td>Tetanus increases as NMR increases</td>
<td>Tetanus increases as NMR increases</td>
</tr>
<tr>
<td></td>
<td>Institutional delivery/skilled attendance rate</td>
<td>Tetanus decreases as skilled attendance increases</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Female literacy rate</td>
<td>Tetanus decreases as literacy increases</td>
<td></td>
</tr>
<tr>
<td>Other:asphyxia</td>
<td>Low birth weight rate</td>
<td>Other increases as LBW rate decreases</td>
<td></td>
</tr>
</tbody>
</table>
known cause. Few data from China were identified, despite searching the Chinese language literature and contacting Chinese experts directly.

The proportion of deaths with unknown cause ranged from 0 to 23%, with a median of 2%. The number of deaths with known cause per study ranged from 21 to 3638 (median 5102.5). NMRs ranged from 8 to 89 per 1000 livebirths. Communication with authors was important in increasing information regarding cause of death. Some publications did not mention neonatal tetanus or diarrhoea but authors provided additional information regarding these causes. Even after communication with authors, 19 studies lacked data on one of our selected causes of death (11 diarrhoea, 4 congenital abnormalities, 3 tetanus, and 1 preterm). Two studies lacked information on two causes (congenital and tetanus; congenital and diarrhoea). Asphyxia was recorded in all the studies and, therefore, chosen as the baseline cause for modelling.

There was substantial variation in the distribution of the different causes of death across the studies, especially for the preterm and infection categories (Figure 2b). Many of the studies with high proportions of neonatal deaths due to congenital abnormalities were from populations with a high prevalence of consanguinity.

Model results
The parameter estimates from the multinomial VR model are shown in Table 4. The model explained some of the variation between countries in the congenital abnormalities:preterm and infection:preterm ratios, but explained little or none of the variation in the ratios of asphyxia and ‘other’ to preterm deaths. The parameter estimates from the multinomial model of the study data are shown in Table 4. The model performed quite well in explaining variation in the infection:asphyxia and tetanus:asphyxia ratios and explained some of the variation in the congenital:asphyxia and diarrhoea:asphyxia ratios. The model explained little or none of the variations in the ratios preterm:asphyxia and other:asphyxia.

The results of jackknife analyses of both models are shown in Table 5. For the VR model, the mean observed and predicted proportions were close in both absolute and relative terms (maximum absolute difference 0.7%, maximum relative difference 7%). Differences were slightly larger for the study data model (maximum absolute difference 2.1% (asphyxia), maximum relative difference 21% (diarrhoea)).

Estimates of the distribution of causes of neonatal deaths
The estimated regional distribution of the causes of neonatal deaths is shown in Figure 3, with global point estimates and uncertainty ranges in Table 5. Three major cause groups predominate—preterm birth, birth asphyxia, and infections (sepsis pneumonia, diarrhoea, and tetanus)—with each responsible for approximately one-quarter to one-third of all neonatal deaths. The remaining deaths (approximately half a
million) are distributed across the remaining causes (congenital and ‘other’). There is considerable variation in proportionate cause by region, particularly for neonatal tetanus, diarrhoea, and congenital malformations. A comparison of our country-level estimates for neonatal tetanus deaths with those produced by WHO Vaccines and Biologicals Department shows reasonable agreement; 7 of the 10 countries with the highest numbers of neonatal tetanus deaths according to WHO are in agreement with these predictions.

Discussion
To our knowledge, this is the first set of global estimates for multiple causes of neonatal death, providing uncertainty estimates and detailing inputs and methods. The WHO has used these estimates in the WHR 2005. Our estimates are broadly consistent with the available single cause estimates. Using different approaches, deaths in the year 2000 have been estimated at 220,000 for neonatal tetanus, at 940,000 for asphyxia, and at 1.33 million for prematurity, although the latter includes deaths attributed to preterm birth up to the age of 5 years. Each of these estimates lies well within our uncertainty range for that cause. Simultaneous estimation of all major causes of deaths in a given time period is challenging, limited by a shortage of high-quality input data and by the complex statistical approaches required. Our uncertainty estimates are wide but still do not capture all the potential sources of uncertainty.

Figure 2 Box plots showing the proportional distribution of causes of neonatal mortality for the two different data sources (a) Vital Registration data (44 countries) and (b) Study data (56 studies)
Geographical variation in the causes of death is striking, emphasizing the need for local data for decision-making. The level of NMR is associated with cause-of-death variation—at NMRs of over 45 per 1000 more than half of neonatal deaths are due to infections and tetanus.2

Input data

The estimates for 95% of neonatal deaths (111 countries 3.8 million deaths, study data model) were based on data on fewer than 14,000 neonatal deaths from 56 studies. We excluded many health-facility-based studies because the distribution of causes of death in these studies may not reflect the distribution of causes of death in the general population and the direction of selection bias is not predictable. For example if obstetric referral is effective, then birth asphyxia will be over-represented in facility-based data.73 Conversely, in isolated areas with low demand for facility-based care, facility-based data may under-estimate asphyxia as a cause of death.74 The exclusion of studies with few deaths of known cause (<20) or those reporting less than five causes of neonatal deaths or which had 25% or more unknown cause of death further restricted the data available (Table 1).

Perhaps more importantly, we were unable to identify useable data for many of the world’s poorest countries, which together account for about one-third of neonatal deaths. It is possible that some publications or unpublished data were missed due to language barriers, despite not limiting searches by language. Extensive attempts were made to contact researchers in China, Latin America, and Francophone West Africa. Approximately one-third of the studies included are from India, which accounts for 28% of the world’s neonatal deaths. Data are particularly lacking from central and north-western Africa, central Asia, and much of China.

There was substantial variation in the distributions of causes between individual data sources in both datasets (VR and study data) (Figure 2). Our models explain only

<table>
<thead>
<tr>
<th>Ratio</th>
<th>Explanatory variable</th>
<th>$R^2$ a</th>
<th>Parameter estimate</th>
<th>95% CI b</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Vital Registration data</td>
<td>Infection: Preterm</td>
<td>GDP (1000s of US$)</td>
<td>0.41</td>
<td>-0.141</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GDP squared</td>
<td>0.0024</td>
<td>0.0018–0.0030</td>
</tr>
<tr>
<td></td>
<td>Congenital: Preterm</td>
<td>Low birth weight rate (%)</td>
<td>0.46</td>
<td>-0.132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Country in EMRO c</td>
<td>1.678</td>
<td>1.296–2.060</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female literacy rate (%)</td>
<td>0.042</td>
<td>0.017–0.066</td>
</tr>
<tr>
<td></td>
<td>Asphyxia: Preterm</td>
<td>Low birth weight rate (%)</td>
<td>0.09</td>
<td>-0.098</td>
</tr>
<tr>
<td></td>
<td>Other: Preterm</td>
<td>None</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>(b) Study data (56 studies)</td>
<td>Infection: asphyxia</td>
<td>BCG coverage (%)</td>
<td>0.57</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Neonatal mortality rate per 1000 livebirths</td>
<td>0.010</td>
<td>-0.001 to 0.020</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female literacy rate (%)</td>
<td>-0.009</td>
<td>-0.016 to -0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Study of early neonatal deaths only</td>
<td>-0.716</td>
<td>-1.080 to -0.351</td>
</tr>
<tr>
<td></td>
<td>Tetanus: asphyxia</td>
<td>NMR per 1000 livebirths</td>
<td>0.55</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female literacy rate (%)</td>
<td>-0.017</td>
<td>-0.037 to 0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Antenatal tetanus toxoid coverage (%)</td>
<td>-0.015</td>
<td>-0.034 to 0.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Study of early neonatal deaths only</td>
<td>-1.743</td>
<td>-2.616 to -0.870</td>
</tr>
<tr>
<td></td>
<td>Diarrhoea: asphyxia</td>
<td>Neonatal mortality rate per 1000 livebirths</td>
<td>0.25</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Study of early neonatal deaths only</td>
<td>-1.145</td>
<td>-2.573 to 0.028</td>
</tr>
<tr>
<td></td>
<td>Congenital: asphyxia</td>
<td>Neonatal mortality rate per 1000 livebirths</td>
<td>0.27</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>% Of institutional deliveries</td>
<td>0.011</td>
<td>0.003–0.018</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Country in EMRO</td>
<td>0.670</td>
<td>0.303–1.037</td>
</tr>
<tr>
<td></td>
<td>Preterm: asphyxia</td>
<td>% Of skilled attendance</td>
<td>0.14</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low birth weight rate (%)</td>
<td>0.025</td>
<td>0.007–0.044</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Study distinguished preterm and term small for gestational age infants</td>
<td>0.289</td>
<td>-0.116 to 0.695</td>
</tr>
<tr>
<td></td>
<td>Other: asphyxia</td>
<td>Study of early neonatal deaths only</td>
<td>0.05</td>
<td>-0.683</td>
</tr>
</tbody>
</table>

a $R^2$-value obtained when fitting the log(ratio) using linear regression with each study having equal weight.

b Estimated using robust standard errors adjusting for within country correlations.

c The majority of countries in the EMRO region have relatively high proportions of consanguinity.

The exclusion of studies with few deaths of known cause (<20) or those reporting less than five causes of neonatal deaths or which had 25% or more unknown cause of death further restricted the data available (Table 1).
some of this variability, although an inability to predict ratios involving the ‘other’ cause category is not surprising. Some of the variation in proportionate mortality by cause shown in the input data is likely to be due to true epidemiological variation; for example, in Figure 2 the outlying studies with a higher proportion of tetanus deaths were from populations with extremely low (<10%) tetanus immunization coverage. However inconsistencies in the attribution of cause of death may also play an important role. Attributing each death to a single cause is an oversimplification. Preterm birth is both a direct cause of death and also a risk factor for other specific causes, notably infections. Some conditions may be synergistic, for example infection and asphyxia.

The variability observed in the VR data was less than that in the study data. The VR data using detailed 4-digit codes allow more specific diagnosis; for example there are multiple specific complications of preterm birth defined rather than a single category of prematurity. Nevertheless, preferences were apparent for certain codes in certain countries. Community-based studies frequently utilize verbal autopsy (VA) approaches, whereby an interviewer gathers information...
regarding the death and a single cause is assigned. VA methods vary from a non-structured interview to detailed post-mortem questionnaires with computer algorithms or several experts assigning a cause of death. The numbers of causes of neonatal death also vary between tools, from four simple groupings to multiple specific diagnoses.

The lack of consistent case definitions and rules in the hierarchical assignment of causes hinders comparisons across time and between studies, and particularly between VR and VA data. Misclassification between causes of neonatal death is not well studied and may particularly affect the infection and preterm categories. Congenital abnormalities, especially cardiac defects, are often missed, especially in VA tools. Improved tools with explicit hierarchies, linking VA and VR data, and with known performance characteristics are required. Two of the studies included here used a VA tool that mapped onto ICD categories, an approach worth further study.

Modelling

The modelling approach used builds on that used previously for child deaths, based on Seemingly Unrelated Regression applied to log ratios of causes. We used multinomial regression models and believe this offers a number of advantages. First, with some assumptions about the category into which unreported causes have been assigned, this approach can handle studies that do not provide information on all the causes of death being modelled. Using the log ratio approach, such studies were excluded. Second, the log ratio approach faces a problem with rarer causes that result in zero deaths in a proportion of data sources. A non-zero value must be introduced, but the choice of which non-zero value to use may affect the results obtained from the model. A jackknife analysis suggested that the log ratio approach underestimated the rarest cause of death in the under-5 analysis, measles. The multinomial model models zeros naturally. A further difference between the two models is in the default weights they give to observations. The log ratio approach, by default, gives equal weight to each study, regardless of size. The multinomial model, by default, gives equal weight to each death, attributing too much weight to large studies when there is within study correlation. We, therefore, chose an intermediate weighting in which each death in a given study carried a weight equal to \( 1/\sqrt{N} \) where \( N \) was the number of deaths included in that study, although the choice of weights made little difference to the model outputs.

Conclusions

To prevent 4 million neonatal deaths we need to know what is causing them. This exercise has highlighted the paucity of reliable, representative data on the causes of neonatal death from settings in which most neonatal deaths occur. Complex statistical models are not a panacea. Counting births and deaths, refining cause-of-death attribution tools and strengthening national child health epidemiology skills all require systematic attention. Each newborn has a right to be counted, and each death should count to prevent others.

Acknowledgements

This work was conducted in collaboration with the Child health Epidemiology Group (CHERG) coordinated by the Department of Child and Adolescent Health and Development of the World Health Organization and with financial support from the Bill & Melinda Gates Foundation. The CHERG leadership including Jennifer Bryce and Cynthia Boschi-Pinto (WHO) are thanked for initiating this work. We are grateful to other CHERG members for constructive criticism, especially Alex Rowe (Centres for Disease Control, Atlanta) for detailed suggestions, and to Melissa Thumm and the abstraction team at the London School of Hygiene and Tropical Medicine (LHTM), particularly Chiedza Zigoni. We thank Igor Rudan, Ozren Polasek, and Ivana Kolic, of the Department of Medical Statistics, Epidemiology, and Medical Informatics at the School of Public Health ‘Andrija Stampar’, Faculty of Medicine of the University of Zagreb in Zagreb, Croatia for their work in preparing Figure 3. We thank the investigators who shared unpublished datasets (marked with asterix) or supplemented their published data including the following: Dr A Aguilar (BASICS II, Bolivia), Dr J Aleman (University of Leon, Nicaragua), Prof K Anand (Pondicherry Medical Institute, Haryana, India), Dr A Bang (SEARCH, Gadchiroli, India), Prof F Barros (formerly of Universidade Federal de Pelotas, Brasil), Prof Z Bhutta* (Aga Khan University, Karachi), Dr V Chongsuvivatwong (Prince of Songkla University, Hat Yai, Thailand), Prof A Dawodu (Faculty of Medicine, University of the United Arab Emirates), Dr J Dommesse (retired University of Cape Town, South Africa), Dr E Ekanem (formerly Department of Paediatrics, University of Calabar, Nigeria), Dr AM El-Shafei (Arabian Gulf University College of Medicine and Medical Sciences, Bahrain), Dr V Fauveau (formerly ICCDR, Bangladesh), Dr F Fikree (The Population Council, Pakistan), Dr P Fonseka (Faculty of Medicine, Galle, Sri Lanka), Dr S Gupta (Medical College, Jaipur, India), Dr S Horpaopan (Children’s Hospital, Rajvithi Hospital, Bangkok, Thailand), Dr A Greenwood (formerly of The MRC Unit, The Gambia), Prof N Khalique (Medical College, Alligarh, India), Dr A Leach (formerly of the MRC Unit, The Gambia), Prof H Perry* (formerly Hospital Albert Schweitzer, Haiti), Prof J Mahanta (Regional Medical Research Centre, Dibrugarh, India), Prof A Pratinidhi (Dept of Prevention and Social Med, Pune, India), Dr N Raina (formerly Dr Datta of The Postgraduate Institute of Medical Education and Research, Chandigarh, India), Dr M Samms-Vaughn (University of West Indies, Kingston, Jamaica), Dr P Settel and the team of the Adult Mortality and Morbidity Project, Tanzania*, Dr S Soemantri (National Institute for Health Research and Development, Indonesia) along with Dr C Surjadiaga for Bahasi translation, Dr E Swedberg (Save the Children, Westport, USA), Dr G Walraven* (formerly of the MRC Farafena Center, The Gambia), Prof D Woods* (retired University of Cape Town, South Africa), and Dr S Zaman (King Edward Medical College, Lahore, Pakistan).

Conflict of Interest

None.
**KEY MESSAGES**

What is known already:
- There are an estimated 4 million neonatal deaths each year.
What this study adds:
- Only ~2.5% of neonatal deaths have reliable cause-of-death information available through vital registration systems. Systematic estimates of the distribution of causes for the remaining 97.5% of deaths are important to guide intervention and funding priorities.
- The major direct causes of neonatal deaths at global level are infections including tetanus (estimated proportion, 35%), preterm birth (28%), and birth asphyxia (23%). There is geographical variation in the proportionate cause of death.
- The substantial uncertainty around these estimates is inevitable given the limited quantity and quality of data from the settings in which the great majority of neonatal deaths occur. More data with consistent attribution of causes are required.
- Several studies included here are based on sample registration systems and use verbal autopsy tools that map cause of death onto ICD codes, a potentially promising approach for scaling up cause-of-death information for the poor.

**References**


Commentary: Utilizing information on causes of neonatal deaths in less-developed countries

H P S Sachdev

Global estimates indicate that the neonatal component of deaths in children younger than 5 years (38% in 2000) is increasing. Although there are substantial regional variations in these proportions (24% in Africa to 56% in the western Pacific), even in areas with similar neonatal mortality rates (47% in Southeast Asia and 26% in sub-Saharan Africa), it is obvious that the Millennium Development Goal for child survival cannot be met without substantial reductions in neonatal mortality. For effective neonatal survival efforts, reasonably accurate information about the causes of deaths is essential.

Using rigorous and transparent methods, Lawn et al. have provided, for 193 countries in the year 2000, systematic estimates, with associated uncertainty, of the distribution of neonatal deaths for programme-relevant causes. Only 2.5% of neonatal deaths had reliable cause-of-death information available through vital registration systems, primarily in well-developed countries. The distribution of reported causes of deaths varied substantially between countries and across studies. The major direct causes of neonatal deaths globally were estimated to be infections (sepsis, pneumonia, tetanus, and diarrhoea; 35%), preterm birth (28%), and birth asphyxia (23%). There was substantial uncertainty around these estimates owing to the limited quantity and quality of data from settings in which the great majority (97.5%) of neonatal deaths had occurred. It is obvious that complex statistical models are not a panacea, and collecting reliable information on the causes of deaths should receive due attention. Simultaneously, it would be pertinent to highlight other important limitations and practical implications of this review.

Several studies included in this review, primarily from low-income and middle-income countries, used verbal autopsy (lay reporting) tools to determine the causes of deaths. The neonate, especially in the first week of life, offers a limited symptomatic response to different insults resulting in substantial overlap of clinical manifestations of several diseases. The sensitivity and specificity of verbal autopsy technique in correctly identifying various causes of deaths is, therefore, sub-optimal. Validation studies suggest that although major congenital malformations and neonatal tetanus have high levels of sensitivity (0.79–0.98 for tetanus) and specificity (≥0.9), the performance for other important conditions like sepsicaemia is poor. The instrument using both open-ended and closed questions identified at least one diagnosis accurately in 71% of the newborns in Pakistan, and resulted in a best agreement (kappa index) of 0.64 for all causes of deaths in India. Comparisons of verbal autopsy data collected and/or analysed by paramedical personnel, paediatricians, and computer algorithms yield disparate results for some but not all causes. Factoring for these aspects would have increased the uncertainty of the estimates provided by Lawn et al. Thus verbal autopsy data can only be a crude pointer to the leading causes of deaths in neonates with understandable limitations for health impact evaluation.

Customarily each death is attributed to a single cause, which is an oversimplification. Clinicians would, however, concur that neonatal deaths invariably have multiple and competing causes. In an urban hospital in New Delhi, India, 75.2% of 129 subjects between the ages of 1 week and 2 months had two or more co-existent morbidities; the co-morbidities being higher in children with a relatively severe condition (2.3 vs 1.8; P = 0.002). Only recently, attention has begun to be directed to methodological advances to account for co-morbidity at the time of death. Appropriate translation into macro-level estimates, and its subsequent adoption for programmatic interventions, therefore, seems distant. Intuitively, what are the possible

Division of Pediatrics and Clinical Epidemiology, Sitaram Bhartia Institute of Science and Research, B-16 Qutab Institutional Area, New Delhi 110 016, India.

Correspondence to: E-6/12 Vasant Vihar, New Delhi 110057, India. E-mail: hpsachdev@gmail.com

Published by Oxford University Press on behalf of the International Epidemiological Association
© The Author 2006; all rights reserved. Advance Access publication 27 April 2006
International Journal of Epidemiology 2006;35:718–719
doi:10.1093/ije/dyl077