clinmedres.org

Using Historical Vital Statistics to Predict the Distribution of Under-Five Mortality by Cause

Chalapati Rao, MBBS, MPH, PhD; Timothy Adair, PhD; and Yohannes Kinfu, PhD, MPhil, MA, BEcon

Background: Cause-specific mortality data is essential for planning intervention programs to reduce mortality in the under age five years population (under-five). However, there is a critical paucity of such information for most of the developing world, particularly where progress towards the United Nations Millennium Development Goal 4 (MDG 4) has been slow. This paper presents a predictive cause of death model for under-five mortality based on historical vital statistics and discusses the utility of the model in generating information that could accelerate progress towards MDG 4.

Methods: Over 1400 country years of vital statistics from 34 countries collected over a period of nearly a century were analyzed to develop relationships between levels of under-five mortality, related mortality ratios, and proportionate mortality from four cause groups: perinatal conditions; diarrhea and lower respiratory infections; congenital anomalies; and all other causes of death. A system of multiple equations with cross-equation parameter restrictions and correlated error terms was developed to predict proportionate mortality by cause based on given measures of under-five mortality. The strength of the predictive model was tested through internal and external cross-validation techniques. Modeled cause-specific mortality estimates for major regions in Africa, Asia, Central America, and South America are presented to illustrate its application across a range of under-five mortality rates.

Results: Consistent and plausible trends and relationships are observed from historical data. High mortality rates are associated with increased proportions of deaths from diarrhea and lower respiratory infections. Perinatal conditions assume importance as a proportionate cause at under-five mortality rates below 60 per 1000 live births. Internal and external validation confirms strength and consistency of the predictive model. Model application at regional level demonstrates heterogeneity and non-linearity in cause-composition arising from the range of under-five mortality rates and related mortality ratios.

Conclusions: Historical analyses suggest that under-five mortality transitions are associated with significant changes in cause of death composition. Sub-national differentials in under-five mortality rates could require intervention programs targeted to address specific cause distributions. The predictive model could, therefore, help set broad priorities for interventions at the local level based on periodic under-five mortality measurement. Given current resource constraints, such priority setting mechanisms are essential to accelerate reductions in under-five mortality.

Keywords: Causes of death; Predictive model; Under-five mortality; Vital registration

Improvements in child survival and health are universally accepted health system goals. The United Nations has called for a two-thirds reduction in global mortality rates in the under age five years population (under-five) over the period 1990 through 2015. Reliable data on total under-five mortality rates (ie, from all causes together) at national and sub-national levels are essential for monitoring progress toward this target. In addition, related information on cause of death is necessary epidemiological data for appropriate interventions to reduce under-five mortality. Efficient civil registration and vital statistics systems that routinely capture all deaths with medically-certified causes are the 'gold standard' source for total and cause-specific mortality data. However, such systems are lacking in many developing countries, leading to a critical paucity of information on local data on levels and causes of under-five mortality, particularly for those populations where such measures are required.²

In the absence of local vital statistics in developing countries, periodic household surveys have been used to estimate total under-five mortality rates. Surveys that estimate total underfive mortality rates include the Demographic Health Survey program operated by Macro International,³ and the Multiple Indicator Cluster Survey program operated by the United Nations Children's Fund. Despite the potential for recall bias and uncertainty regarding reference time periods for surveybased estimates,5 total under-five mortality rates derived from these sources have been used for tracking global and national trends in under-five mortality.⁶⁻⁸ However, small sample sizes only permit total mortality estimates at national or major subnational levels, and long intervals between surveys hinders timely health policy responses to these estimates. Also, these surveys do not routinely collect information on causes of death that are principal inputs for planning health interventions and are, therefore, only useful at best for monitoring under-five mortality, but not for informing targeted health service delivery.

In the absence of information on causes of death, statistical models have been developed to predict cause-specific mortality among people under five years of age using total under-five mortality rates as model inputs. 9,10 In this paper, we present a statistical prediction model based on historical relationships between levels of total under-five mortality and proportionate mortality from four cause groups, as observed in vital registration data from 34 countries over a period of approximately a century (1400 country-years of data). The model consists of a system of regression equations that predict proportionate mortality by cause, using given levels of under-five mortality and related mortality ratios as model inputs. We also present the application of the statistical model using a range of input variables to predict cause-specific proportionate mortality for broad geographical regions of the world. The results serve as an example of the utility of the model in planning clinical intervention packages for

populations with varying total under-five mortality patterns, for which cause-specific mortality data is not available.

Methods

Evolution of Cause-of-Death Prediction Models

Historical relationships between levels of mortality and causes of death have been studied extensively in an international context for nearly four decades, starting with Omran's theory of the epidemiological transition. 11 Subsequently, Preston¹² quantified the relationship in a set of regression equations fitted to historical vital registration data from Western countries. Since then, vital registration data have been used to develop predictive models to estimate cause-specific mortality for countries where data are inadequate or unavailable. 9,13-14 Salomon and Murray 15 introduced the concept of "compositional models" to ensure internal consistency, using proportionate mortality by cause as the variable of interest, rather than cause-specific mortality rates. These models were extensively used in the Global Burden of Disease Project to estimate cause-specific mortality across all ages for several countries in 2001.16

Design and Limitations of Previous Cause-of-Death Models for Under-Fives

To meet a specific need for cause-specific mortality estimates at ages below five years, Morris et al 10 designed compositional models to predict cause of death patterns among under-fives for countries in Asia and Africa in the year 2000. Their models were built using data on levels and cause distributions of under-five mortality from 46 selected research studies conducted in developing countries during 1980-1995. All these studies employed verbal autopsy methods to ascertain causes of death. The models were developed to predict proportionate mortality from six causes: diarrheal diseases, pneumonia, malaria, measles, neonatal and other causes, and undetermined causes. The authors cited several potential limitations of their approach, including variations in research design, data collection methods, and classification of causes of death across studies. Further, Morris et al10 estimated their predictive model as a function of several explanatory variables (eg. access to safe water, immunization, and skilled birth attendance) in addition to the under-five mortality rate. For each population represented in the research studies used in the model, data on these additional explanatory variables were pooled from other external sources and were subject to respective measurement errors. Another limitation noted was that their models did not predict proportionate mortality from perinatal conditions, which is required for planning newborn care services.17

New Design of Cause-of-Death Models for Under-Fives In order to overcome some of the limitations in the previous models, we adopted an approach using different data sources and analytical processes. Firstly, we chose to use all available historical vital registration data up to 2008 to explore the relationships between levels and causes of under-five mortality, and to use these data for developing the models. We

Table 1. Cause group-specific codes from various revisions of the ICD.

Cause group	ICD-2	ICD-3	ICD-4	ICD-5	ICD-6/7	ICD-8	ICD-9	ICD-10
Diarrhea and LRI	1, 10, 12-14, 89,	1a-b, 11a-b, 14-	1, 2a-b, 11a-b,	1, 2a-c, 4, 27a-c,	A012, A014-	A001, A002,	001, 002,	A00, A01,
	91-92, 104-105,	15, 16a-c, 99a,	12, 13a-c,	33a-b, 106a, 107-	A016, A104,	A004, A005,	004, 006-	A03, A04,
	113-114	100, 101a-b,	106a, 107-	109, 117b, 119,	A088-A092	A090-A092	009, 466,	A06-A09,
		111b, 120a-b	109, 117b,	123a, 572			480-487	J10-J18,
			119, 120a-b					J20-22
Perinatal	151-153	160, 161a-b,	158-159,	158-159, 160a1-	A130-A135	A131-A135	760-779	P00-P96
		162-163	160a-b, 161a-c	b2, 161a-e				
Congenital anomalies	150	159	157a-c	157a-j	A127-A129	A126-A130	740-759	Q00-Q99

LRI, Lower respiratory infections.

chose to use vital registration data in view of the intrinsic strengths of these data in terms of the following:

- complete (or near complete) population coverage, and therefore larger sample sizes;
- uniformity in death reporting procedures under legal provisions, with medical certification of cause of death wherever feasible;
- consistency in cause of death classification across countries, according to prevalent versions of the International Classification of Diseases (ICD).¹⁸

Consistent time series for specific causes of death are constrained by revisions to ICD lists over time as well as differences in the levels of aggregation by cause for different country years of data. Hence, we restricted our analysis to four major mutually exclusive cause groups, the choice of which was also guided by our interest in providing estimates that would inform the development of national and local intervention programs. The ICD codes used in this study are listed in table 1. The four cause groups used are as follows:

- a combined cause group of deaths from diarrheal diseases and lower respiratory infections (LRI) to simulate the lower limit of proportion of deaths from infectious diseases that are the focus of Integrated Management of Childhood Illness (IMCI) intervention package;
- deaths from conditions originating in the perinatal period (mainly prematurity/low birth weight, birth asphyxia, and birth trauma among others, but excluding stillbirths) as an estimate of the proportion of deaths that could be addressed through the provision of adequate newborn care:
- deaths from congenital malformations to accurately capture the proportion of largely unpreventable deaths;
- deaths from all other causes including vaccine preventable diseases, meningitis, malaria, injuries, childhood cancers, ill-defined or unknown causes, and all other specified conditions.

Data Sources

The World Health Organization mortality database was used as the primary source of vital registration data on age-specific population and cause-specific mortality by sex from countries reporting these statistics since 1950.¹⁹ Additionally, vital registration data from certain countries for the period prior to 1950 was sourced from a compilation used by Preston et al¹² in previous analyses. Criteria used to include country specific data for this analysis were that time series data spanning at least three decades were available and were adjudged to be of adequate quality for direct estimation of cause-specific mortality for the Global Burden of Disease 2001 Study.¹⁶ Table 2 shows the country-years of data used in the analysis.

For each country-year, the following variables were computed:

- under-five mortality rate (5q0) by sex;
- ratio between infant mortality rate (IMR or 1q0) and underfive mortality rate (1q0/5q0);
- ratio between child mortality rate (CMR or 4q1) and underfive mortality rate (4q1/5q0);
- male-to-female ratio of under-five mortality rate;
- ratio between neonatal mortality rate (NMR) and underfive mortality rate (NMR/₅q₀) (based on data availability as shown in table 2); and
- proportionate mortality by cause according to the four cause groups of interest in this study.

Descriptive analyses were conducted to observe trends in the relationships between these variables with declines in underfive mortality. Subsequently, following Salomon and Murray¹⁵, and Morris¹⁰ and others, we then built a compositional model of the following form:

$$\ln(P_{ii} / P_{iC}) = \beta_0 + \beta_1 \ln(_5q_0) + \beta_2 Q_{i1} + \beta_3 Q_{i2} + \beta_4 Q_{i3} + e_{ii}$$

where P_{ij} is the cause proportion for cause j (where j=1, 2, 3 and represent perinatal, diarrhea-LRI, or 'other'), P_{iC} is cause proportion for congenital anomalies (the base cause), ${}_5q_0$ is the under-five mortality rate, Q_{i1} is ${}_1q_0/{}_5q_0$, Q_{i2} is ${}_4q_1/{}_5q_0$, Q_{i3} is ${}_5q_0$ (male)/ ${}_5q_0$ (female), β is the parameter to be estimated, e_{ij} is the normally distributed error term, and i is each observation. For the compositional model, three linear equations (for each cause j) were solved simultaneously using Seemingly Unrelated Regression, 20 for which the coefficients are

presented in the results. Also, the r^2 for each model is presented, which measures the proportion of variation in the outcome variable that is accounted for by the regression.²¹

The strength of the predictive model was tested through internal and external cross-validation techniques. Internal cross-validation was conducted in two ways. Firstly, we applied the k-fold technique to the observed data (Appendix 1).²² Second, we compared the observed cause proportions with the model predicted values for the complete dataset; and also for a sample of country years with 5q0 above 60 per 1000. For external cross-validation, we compared predicted cause proportions with proportions observed in vital registration data from South Africa for 2005. The root mean square error (RMSE) is used to measure the magnitude of error between the observed and predicted cause proportions.

We also applied the model to estimate cause-specific proportionate mortality among under-fives in major subdivisions of Africa, Asia, and parts of the Americas (Appendix 2), where little empirical evidence of this nature is available. Regional values of model input variables for the period 2005 through 2010 were extracted from the United Nations World Population Prospects Database.²³ All analyses were undertaken using Stata 11.0 (Stata Corporation, TX).

Results

The complete dataset represents 1400 country-years of data. The mean ${}_5q_0$ from the full dataset is 33 per 1000 for males (ranging from 4 to 362), and 27 per 1000 for females (ranging from 4 to 338). Over 10% of country-years have a ${}_5q_0$ of over 70 per 1000, which are similar to current global averages and to levels currently prevalent in more than a third of the world's population residing largely in Africa and southern Asia.

Table 2. Country-years of data used in historical analyses and models for predicting cause-composition of under-five mortality.

Country	Years used in predictive model	Years for which neonatal mortality data available
Argentina	1968 to 1970, 1977 to 1996	1968 to 1970, 1977 to 1979, 1982 to 1996
Australia	1911, 1921, 1933, 1940, 1947, 1950 to 1996	1968 to 1996
Austria	1955 to 1999	1969 to 1999
Belgium	1954 to 1994	1968 to 1992
Bulgaria	1964 to 1999	-
Canada	1921, 1931, 1941, 1950 to 1996	1969, 1971, 1986 to 1996
Chile	1930, 1940, 1955 to 1993	1968 to 1982, 1986 to 1993
Costa Rica	1961 to 1995	1975 to 1976, 1980, 1982 to 1995
Denmark	1930, 1940, 1951 to 1992	1968 to 1976, 1978 to 1992
England/Wales	1911, 1921, 1932, 1941, 1951 to 1957,	1968 to 1998
9	1963 to 1966, 1968 to 1998	
Finland	1952 to 1995	1969 to 1995
France	1950, 1952 to 1997	1968 to 1997
Greece	1956 to 1997	1968 to 1997
Hong Kong	1960 to 1999	1972 to 1975, 1985 to 1999
Hungary	1965 to 1998	1969 to 1998
Ireland	1950 to 1998	1976 to 1998
Israel	1977 to 1999	1980 to 1999
Italy	1931, 1951 to 1997	1968 to 1997
Japan	1950 to 1999	=
Mexico*	1958 to 1967, 1969 to 1976, 1978 to	1969 to 1976, 1979 to 1980, 1989 to
	1983, 1987 to 1995	1990, 1992 to 1993
Netherlands	1931, 1940, 1950 to 1999	1969, 1971 to 1990, 1998, 1999
New Zealand	1911, 1921, 1926, 1936, 1945, 1950 to 1998	1968 to 1998
Norway	1946, 1962 to 1965, 1971, 1975 to 1998	1971 to 1998
Poland	1960 to 1995	1969 to 1995
Portugal	1955 to 1999	1971 to 1999
Romania	1959 to 1999	-
Singapore	1963 to 1998	1968 to 1998
Spain	1951 to 1969, 1971 to 1991, 1993 to 1998	1975 to 1998
Sweden	1951 to 1996	1969 to 1996
Switzerland	1951 to 1994	1969 to 1994
	1951 to 1963, 1966 to 1968, 1970 to 1994	-
Uruguay	1955 to 1960, 1963 to 1970 to 1990	1970 to 1972, 1977
USA	1910, 1930, 1940, 1950 to 1960, 1964 to 1998	1968 to 1998
Venezuela	1955 to 1983, 1985 to 1990, 1992 to 1994	1968, 1970 to 1972, 1977, 1985 to 1994

^{*1699} to 1971, no date for females.

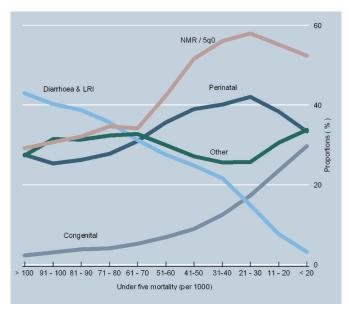


Figure 1. Cause proportions and the ratio of the neonatal mortality rate to under-five mortality rate at different levels of under-five mortality rate from historical vital registration data.

Figure 1 shows the secular changes in cause-composition of under-five mortality with declining mortality rates. Graphed variables are averaged across deciles of under-five mortality rates. The graph depicts a consistently positive relationship between the proportion of deaths due to diarrhea and LRI and under-five mortality rates, and a consistently negative relationship for the proportion of deaths due to congenital anomalies. Also, the proportion of deaths due to perinatal conditions increases sharply from 25% to 40% as $_5q_0$ falls below 60 per 1000. This is accompanied by an increasing ratio of NMR/ $_5q_0$ indicating the compression of under-five mortality into the neonatal period as child mortality declines. Finally, the proportion of deaths classified to 'other' conditions

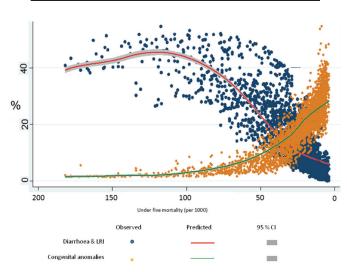


Figure 2. Variability in observed and mean predicted proportionate mortality from diarrhea and LRI and congenital anomalies, at different levels of under-five mortality rates.

varies between 25% to 35% throughout the range of levels of child mortality shown in figure 1. To provide additional perspective, figure 2 depicts the variability in proportionate mortality from the two causes that have consistent relationships with under-five mortality rates (diarrhea and LRI, and congenital anomalies), and the 95% confidence intervals of the mean proportions for each cause across the range of under-five mortality rates represented in the data. This graph indicates that despite the variability, the data support the clear trend lines of the mean proportions for both causes.

Table 3 presents the regression coefficients in the predictive model for each cause group. Interpretation of the results needs to consider that the coefficients represent the value of each cause relative to deaths from congenital anomalies. The r^2 for each of the three cause models is strong, and is highest for the ratio of diarrhea and LRI deaths to congenital deaths ($r^2 = 0.86$). The log of ${}_5q_0$ is positively related to all the dependent variables in the three regression models because of its negative correlation with the base cause 'congenital anomalies'. Each of the other variables in the predictive model is also significantly related to the respective outcome variables.

Table 4 shows the results from the internal and external cross-validation techniques of the predictive model. The internal validation comparing observed and predicted cause proportions for the full sample shows the model is most accurate for congenital malformations (RMSE of 4.3%), and is also reliable for the other three causes (RMSE <10%). For high mortality country years ($_5q_0$ is above 60), the RMSE is very low for congenital malformations and remains reasonable for each of the remaining causes (around 10). Similar correlation is observed from external validation using South African data. The additional internal cross-validation using the k-fold technique (Appendix 1) shows the RMSEs to be consistent across the 10 sub-samples, and is $\leq 10\%$ for each cause. Taken together, these results indicate the predictive accuracy of the model.

The application of our model to predict cause composition of under-five mortality in regions of Africa, Asia, and the Americas reveals the heterogeneity of patterns represented in the historical data for different levels of $_5q_0$ (table 5). For instance, where there are very high levels of 5q0 as in Western and Middle Africa (>150 per 1000), the model predicts 'other causes' to be the leading cause group, which together with 'diarrhea and LRI' account for over 80% of deaths. It is to be noted that the 'other causes' group would include deaths from infectious diseases such as malaria, measles, and meningitis. At high levels (80-150 per 1000), diarrhea and LRI are the leading cause group. These results reflect the flexibility of the model at depicting a non-linear relationship between the level of ₅q₀ and the causes that are associated with relatively high proportion of deaths occurring beyond infancy (ie, over 40%). At these levels of 5q₀, infectious diseases are also likely to be the major cause group during infancy, with the model

Table 3. Regression coefficients for prediction model.

		Standard		
Model parameters	Coefficient	Error	r²	RMSE
Diarrhea and LRI			0.858	0.637
log (₅ q ₀)	2.106	0.027*		
$_{4}q_{1}$ / $_{5}q_{0}$	-41.016	4.064*		
₁ q ₀ / ₅ q ₀	-47.822	4.265*		
$_{5}q_{0}$ (male) / $_{5}q_{0}$ (female)	-1.335	0.141*		
Constant	56.049	4.362*		
Perinatal			0.757	0.359
log (₅ q ₀)	0.682	0.015*		
$_{4}q_{1}$ / $_{5}q_{0}$	21.402	2.292*		
₁ q ₀ / ₅ q ₀	22.059	2.405*		
$_{5}q_{0}$ (male) / $_{5}q_{0}$ (female)	-0.263	0.079*		
Constant	-18.252	2.460*		
Others			0.701	0.440
log (₅ q ₀)	0.543	0.018*		
4q ₁ / ₅ q ₀	24.924	2.809*		
1q ₀ / ₅ q ₀	20.126	2.948*		
$_{5}q_{0}$ (male) / $_{5}q_{0}$ (female)	-0.585	0.097*		
Constant	-17.677	3.015*		

^{*}P value < 0.01

LRI, Lower respiratory infections; RMSE, Root mean square error.

predicting very low proportions of deaths due to perinatal conditions. At intermediate levels (40-70 per 1000), perinatal conditions have increasing importance, although infectious diseases remain a threat, possibly occurring with increasing frequency during the neonatal period, as under five-mortality declines. At lower levels of $_5q_0$, perinatal conditions dominate the cause-specific mortality patterns.

Discussion

Recent analyses of national trends in under-five mortality⁷ and coverage of specific interventions²⁴ indicate slow progress, with even reversal of mortality reduction in some countries. These international assessments have attracted global attention to this unsatisfactory situation. These analyses have also been supplemented by efforts to estimate the cause composition of under-five mortality at global, regional, and even national levels as evidence to support policy formulation.²⁵ The historical analyses presented here guide the understanding of the trajectory of changing composition of under-five mortality across the entire range of mortality levels and provide an empirical basis for planning initiatives to reduce under-five mortality. The predictive models provide information on

proportionate mortality that is required for health sector priority setting mechanisms that take into account the magnitude of specific conditions, the cost-effectiveness of available interventions, and the feasibility in implementing them.²⁶ Indeed, the need for epidemiological evidence on under-five mortality is a growing perception in the health policy environment.²⁷

A potential limitation in our models could be the variability in input data arising from changes to ICD classifications over time. To overcome this limitation, we chose four broad cause categories of interest to minimize any potential for misclassification that could result from ICD revisions. For instance, there have been changes and additions to the codes used for diarrheal diseases based on identification of new causative microorganisms, but these are all included within the broad category for diarrhea across all ICD revisions. Further, by combining diarrhea and LRI into one category, we accounted for potential misclassification between these causes in underlying cause data. We chose to keep measles and other vaccine preventable diseases in the 'other causes' category for two reasons. Firstly, there is considerable variability in

Table 4. Comparison of observed and estimated cause proportions for internal and external validation of the prediction model.

Cause	Full Sample			₅ q ₀ over 60 per 1000**			South Africa Vital Registration, 2005			
	Observed*	Predicted	RMSE	Observed**	Predicted	RMSE	Observed	Predicted	RMSE	
Diarrhea and LRI	14.9%	14.1%	7.5%	37.5%	40.3%	10.2%	38.2%	37.7%	0.48%	
Perinatal	37.1%	37.5%	8.6%	28.1%	28.7%	8.6%	19.8%	27.1%	7.26%	
Congenital	18.6%	18.9%	4.3%	3.7%	3.9%	1.3%	2.2%	5.3%	3.12%	
Others	29.5%	29.5%	8.6%	30.8%	27.1%	10.4%	39.8%	29.9%	9.91%	

^{*2838} observations

^{**307} observations

LRI, Lower respiratory infections; RMSE, Root mean square error.

Table 5. Model-based predictions of cause-composition of under-five mortality by sex for regions in Africa, Asia, and the Americas, 2005 to 2010.

	Mortality per 1000			Cause proportions (%)					
Region	₅ q ₀			Diarrhea and LRI	Perinatal	Congenital	Others		
Central America	25	6	20	15.1	36.3	14.7	33.9		
South America	28	7	21	18.3	34.1	12.7	34.9		
Eastern Asia	28	6	22	26.7	29.8	9.6	33.9		
South-Eastern Asia	37	9	28	24.1	33.2	9.8	32.9		
Western Asia	37	7	30	22.3	37.3	10.7	29.7		
Caribbean	45	11	35	30.2	31.1	7.6	31.1		
Northern Africa	56	15	42	36.8	27.4	5.5	30.4		
Southern Africa	72	24	49	42.9	21.8	3.5	31.8		
South-Central Asia	82	27	56	51	18	2.3	28.7		
East Africa	124	52	76	48.8	15.7	1.2	34.3		
Western Africa	166	76	98	43.6	15.4	0.8	40.2		
Middle Africa	189	87	112	35.8	18.2	0.7	45.3		

LRI, Lower respiratory infections.

assignment of these causes in registration data, possibly arising from their clinical presentation as common infectious diseases, as well as due to changes induced by ICD revisions. Secondly, from a policy perspective, immunization programs are mandatory, therefore, need not be included in any priority setting debate.

Ideally, such predictive models should include representative data from high-mortality populations in recent times and should also cover more detailed causes. However, the lack of reliable data meant that the current analysis could only be based on historical data, mostly from developed nations. We address this limitation by applying internal statistical validation techniques as well as by comparing our results with the only reliable contemporary external data set available from South Africa. Our results have shown that the model we developed is able to predict cause-specific distributions at all levels of mortality within a narrow margin of error. In particular, the close correlation between the observed values in South African data and the predicted values from the historical model justifies its utility for current application, at least for estimating proportionate mortality by broad cause groups.

Another possible limitation in our analyses is the assumption that age at death and cause of death are accurately recorded in national vital registration data from the countries included in our analysis. On close scrutiny, the dataset revealed that no deaths from conditions originating in the perinatal period were assigned beyond infancy. In terms of use of condition-specific codes, deaths from diarrhea and LRI were exclusively coded at ages beyond the perinatal period (the first week of life), with the few deaths from these causes within the first week of life being assigned the relevant codes from the perinatal chapter of the ICD.

Recent analyses have identified considerable improvement in the coverage of routine interventions such as immunization, antenatal care, and micronutrient supplementation; however, there has been little progress in the provision of emergency care to detect and cure childhood infectious diseases or deliver obstetric and newborn care.²⁸ There are calls for increasing coverage of different intervention packages;²⁹⁻³⁴ however, local limitations in availability of human and technical resources could restrict comprehensive intervention coverage.³⁵ From a policy perspective, the cause of death classification that we have adopted in our analysis could be used to set priorities in choosing between clinical intervention packages for diarrhea and LRI (IMCI), or newborn care interventions for perinatal conditions, depending on prevailing levels of under-five mortality. Keeping this in mind, we propose a basic framework for prioritizing the implementation of different clinical intervention packages as follows:

- at levels of ₅q₀ >100: focus on interventions to prevent, detect and treat diarrhea, pneumonia, and other infectious causes (ie, the IMCI strategy);
- at levels of 5q0 between 60 and 100: intensify efforts to extend IMCI strategies into the neonatal period, including the first week of life;
- at levels of 5q0 below 60: increase focus on skilled newborn care, including the availability of referral services for adequate clinical management.

Currently, available national estimates of under-five mortality for many countries are likely to mask sub-national differentials, information that could improve health services planning and delivery. Simple mortality surveys including the birth history module and a few relevant questions on health care access or health care seeking behavior could resolve this information gap. Such surveys could be mounted through collaboration between local government and academic institutions. The resultant local estimates of under-five mortality rates could be used to predict proportionate mortality by broad cause groups. This could facilitate local planning for interventions, the impact for which could be monitored through repeat

surveys over time. In conclusion, we propose that to accelerate progress towards achieving MDG 4, it is necessary to probe deeper into sub-national and local patterns of under-five mortality and be more focused or specific in intervention delivery, for which the analyses presented here provide a technical basis for doing so.

References

- Millennium Project. Investing in development: A practical plan to achieve the Millennium Development Goals. 2002-2006. New York, NY: United Nations Secretary General and Development Group. 2005. Available at: http://www.unmillenniumproject.org/reports/index.htm. Accessed: September 27, 2009.
- Mahapatra P, Shibuya K, Lopez AD, Coullare F, Notzon FC, Rao C, Szreter S; on behalf of the Monitoring Vital Events (MoVE) writing group. Civil registration systems and vital statistics: successes and missed opportunities. Lancet 2007. Epub ahead of print.
- MEASURE DHS. Demographic and Health Surveys: Quality information to plan, monitor and improve population, health, and nutrition programs. About DHS. 2009. Available at: http://www.measuredhs.com/aboutdhs/. Accessed: September 28, 2009.
- UNICEF. Unite for Children. Statistics and Monitoring: Multiple Indicator Cluster Survey (MICS). 2009. Available at: http://www.unicef.org/statistics/index_24302.html. Accessed: September 27, 2009.
- Preston SH, Guillot M, Heuveline P. Indirect Estimation Methods. In: Demography: Measuring and modeling population processes. Oxford, UK: Wiley-Blackwell; 2001. 224-246.
- Ahmad OB, Lopez AD, Inoue M. The decline in child mortality: a reappraisal. Bull World Health Organ 200;78:1175-1191.
- Murray CJ, Laakso T, Shibuya K, Hill K, Lopez AD. Can we achieve Millennium Development Goal 4? New analysis of country trends and forecasts of under-5 mortality to 2015. Lancet. 2007;370:1040-1054.
- United Nations. Child Mortality since the 1960s: A database for developing countries. United Nations publication, No. E.92. XIII.10. New York, NY: Department for Economic and Social Development; 1992.
- Lopez AD, Hull TH. A note on estimating the cause of death structure in high mortality populations. Popul Bull UN 1983;14:66-70.
- Morris SS, Black RE, Tomaskovic L. Predicting the distribution of under-five deaths by cause in countries without adequate vital registration systems. Int J Epidemiol 2003; 32:1041-1051.
- Omran AR. Epidemiologic transition: a theory of the epidemiology of population change. Milbank Mem Fund Q 1971;49:509-538.
- Preston SH. Mortality patterns in national populations: with special reference to recorded causes of death. New York, NY: Academic Press; 1976.
- Hakulinen T, Hansluwka H, Lopez AD, Nakada T. Global and regional mortality patterns by cause of death in 1980. Int J Epidemiol 1986;15:226-233.
- Bulatao RA, Stephens PW. Global estimates and projections of mortality by cause, 1970-2015. Policy research working papers: Population, Health and Nutrition. Washington, DC: World Bank WPS: 1992. 1007.
- Salomon JA, Murray CJL. The epidemiological transition revisited. Compositional models for causes of death by age and sex. Population and Development Review 2002;28:205-228.

- 16. Mathers CD, Lopez AD, Murray CJL. The Burden of disease and mortality by condition: data, methods, and results for 2001. In: Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJL, eds. Global Burden of Disease and Risk Factors. New York, NY: World Bank and Oxford University Press;2006.
- 17. Lopez AD. Commentary: Estimating the causes of child deaths. Int J Epidemiol 2003;32:1052-1053.
- 18. World Health Organization. History of the development of the ICD. International Statistical Classification of Diseases and Health Related Problems - Tenth Revision. Volume 2: Instruction Manual. Geneva, Switzerland: World Health Organization; 1993. 139-153.
- Health statistics and health information systems. WHO
 Mortality Database: Tables. World Health Organization.
 Available at: http://www.who.int/healthinfo/morttables/en/.
 Accessed: December 1, 2007.
- Greene WH. Simultaneous-Equation Models. Econometric Analysis. New York, NY: Macmillan Publishing Company; 1990. 591-660.
- Kirkwood B. Sterne J. Measurement error: assessment and implications. In: Essential Medical Statistics. Oxford, UK: Wiley-Blackwell; 2003. 429-439.
- Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York, NY: Springer; 2009. 219-260.
- World Population Prospects: The 2006 Revision. United Nations Department of Economic and Social Affairs; Population Division: Population Database. 2006.
- 24. Bryce J, Terreri N, Victora CG, Mason E, Daelmans B, Bhutta ZA, Bustreo F, Songane F, Salama P, Wardlaw T. Countdown to 2015: tracking intervention coverage for child survival. Lancet 2006;368:1067-1076.
- Bryce J, Boschi-Pinto C, Shibuya K, Black RE, WHO Child Health Epidemiology Reference Group. WHO estimates of the causes of death in children. Lancet 2005;365:1147-1152.
- The 10/90 report on health research 1999. World Health Organization. Geneva, Global Forum for Health Research, 1999.
- 27. Victora CG. Measuring progress towards equitable child survival: where are the epidemiologists? Epidemiology 2007:18:669-672.
- Fauveau V, Lawn JE, Mason E, Newby H, Shankar A, Starrs A, Wardlaw T. Countdown to 2015 for maternal, newborn, and child survival: the 2008 report on tracking coverage of interventions. Lancet 2008;371:1247-1258.
- 29. Darmstadt GL, Bhutta ZA, Cousens S, Adam T, Walker N, de Bernis L, Lancet Neonatal Survival Steering Team. Evidencebased, cost-effective interventions: how many newborn babies can we save? Lancet 2005;365:977-988.
- 30. Knippenberg R, Lawn JE, Darmstadt GL, Begkoyian G, Fogstad H, Walelign N, Paul VK, Lancet Neonatal Survival Steering Team. Systematic scaling up of neonatal care in countries. Lancet 2005;365:1087-1098.
- Gwatkin DR. IMCI: what can we learn from an innovation that didn't reach the poor? Bull World Health Organ 2006;84:768.
- 32. Rudan I, El Arifeen S, Black RE, Campbell H. Childhood pneumonia and diarrhoea: setting our priorities right. Lancet Infect Dis 2007;7:56-61.
- 33. Bhutta ZA, Ahmed T, Black RE, Cousens S, Dewey K, Giugliani E, Haider B A, Kirkwood B, Morris SS, Sachdev HP, Shekar M. What works? Interventions for maternal and child undernutrition and survival. Lancet 2008;371:417-440.
- 34. Black RE, Allen LH, Bhutta ZA, Caulfield LE, de Onis M, Ezzati M, Mathers C, Rivera J; Maternal and Child Undernutrition Study Group. Maternal and child undernutrition: global and regional exposures and health consequences. Lancet 2008;371: 243-260.

35. Kinfu Y, Dal Poz MR, Mercer H, Evans DB. The health worker shortage in Africa: are enough physicians and nurses being trained? Bull World Health Organ 2009;87:225-230.

Author Affiliations

Chalapati Rao, MBBS, MPH, PhD*; Timothy Adair, PhD*; Yohannes Kinfu, PhD, MPhil, BEcon*

Appendix I

K-fold Cross-Validation Exercise

In the k-fold cross-validation exercise, the data set is divided into k subsets, and each time, one of the k subsets is used as a validation data set and the other k-1 subsets, known as the training set, is used for model estimation. Hence, every data point gets to be in a validation set exactly once, and gets to be

in the training set for model building k-1 times. Following standard practice, we implemented a 10-fold cross-validation procedure. This involves dividing the data into 10 subsets, and each time, one of the 10 subsets is used as a validation data set and the other 9 subsets as the training set for model estimation. Then, we run ten sets of models using each of the ten possible training sets. On each run, the regression model estimated from the training data is used to predict proportional mortality outcomes for the validation data set, which then was compared with the actual proportional mortality from the same data. The advantage of this method over other statistical validation procedures is that all observations are used for both training and validation, and each observation is used for validation exactly once. Finally, we test the goodness-of-fit of our model using the root mean square error obtained from the difference between predicted and actual values of each of the 10 sub-sets, and the combined data set. The findings from application of the k-fold technique are shown in the table below.

Appendix 1, Table 1. Root mean square error obtained from 10-fold validation sample.

Sub-sample number										
Cause	1	2	3	4	5	6	7	8	9	10
Diarrhea and LRI	7.51	7.75	8.00	7.08	6.51	7.11	7.83	7.53	7.44	7.88
Perinatal conditions	8.52	8.81	9.10	8.28	8.52	8.48	9.09	8.82	8.41	8.58
Congenital anomalies	3.98	4.42	4.27	4.50	4.47	4.39	3.97	4.31	4.22	4.39
Others	8.92	8.62	9.43	8.54	8.75	8.05	8.88	8.69	8.18	8.02

LRI, Lower respiratory infections.

Appendix 2

Regions

Eastern Africa: Burundi, Comoros, Djibouti, Eritrea, Ethiopia, Kenya, Madagascar, Malawi, Mauritius, Mozambique, Réunion, Rwanda, Seychelles, Somalia, Uganda United Republic of Tanzania, Zambia, Zimbabwe.

Middle Africa: Angola, Cameroon, Central African Republic, Chad, Congo, Democratic Republic of the Congo, Equatorial Guinea, Gabon, São Tomé and Príncipe.

Northern Africa: Algeria, Egypt, Libyan Arab Jamahiriya, Morocco, Sudan, Tunisia, Western Sahara.

Southern Africa: Botswana, Lesotho, Namibia, South Africa, Swaziland.

Western Africa: Benin, Burkina Faso, Cape Verde, Côte d'Ivoire, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Mali, Mauritania, Niger, Nigeria, St. Helena, Senegal, Sierra Leone, Togo.

Eastern Asia: China, China Hong Kong SAR, China Macao SAR, Democratic People's Republic of Korea, Japan, Mongolia, Republic of Korea.

South-Central Asia: Afghanistan, Bangladesh, Bhutan, India, Iran (Islamic Republic of), Kazakhstan, Kyrgyzstan, Maldives, Nepal, Pakistan, Sri Lanka, Tajikistan, Turkmenistan, Uzbekistan.

South-Eastern Asia: Brunei Darussalam, Cambodia, Indonesia, Lao People's Democratic Republic, Malaysia, Myanmar, Philippines, Singapore, Timor-Leste, Thailand, Viet Nam.

Western Asia: Armenia, Azerbaijan, Bahrain, Cyprus, Georgia, Iraq, Israel, Jordan, Kuwait, Lebanon, Occupied Palestinian Territory, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, Turkey, United Arab Emirates, Yemen.

Central America: Belize, Costa Rica, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama.

South America: Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Falkland Islands, French Guiana, Guyana, Paraguay, Peru, Suriname, Uruguay, Venezuela.

^{*}School of Population Health, University of Queensland, Herston, Australia