R01 proposal
"Global Age Patterns of Under-Five Mortality"

PI:
Michel Guillot, University of Pennsylvania

Site PIs:
Patrick Gerland, United Nations Population Division
Joanne Katz, Johns Hopkins Bloomberg School of Public Health
Li Liu, Johns Hopkins Bloomberg School of Public Health
Gilles Pison, Museum National d'Histoire Naturelle
Georges Reniers, London School of Hygiene and Tropical Medicine

Submitted to NICHD on 5 February 2016

Project Summary

The Under-5 Mortality Rate (U5MR) is a key and widely-used indicator of child health, but it conceals important information about how this mortality is distributed by age. For better understanding and monitoring of child health, it is critical to examine how the risk of death varies within the 0-5 age range. This includes age breakdowns beyond the standard cut-off points of 28 days (for neonatal mortality) and 1 year (for infant mortality). In many populations, however, the age pattern of under-5 mortality is not well known. Less-developed countries, in particular, lack the high-quality detailed vital registration information necessary for the analysis of such age patterns. Sample surveys collecting retrospective birth histories do not satisfactorily fill this gap, because they are subject to systematic biases that are particularly consequential for estimating age patterns. This makes the need for high-quality information on age patterns of under-5 mortality even more critical, because regularities in these age patterns can be used as a powerful tool for evaluating and correcting data when sources are deficient. The goal of this project is to improve our understanding of age patterns of under-5 mortality by gathering the largest database to date on high-quality global mortality information by detailed age (by days, weeks, months, and years of age) from birth until age 5, by sex. This database, which will cover a wide array of historical and contemporary contexts in both more- and less-developed settings, will serve as a basis for generating models summarizing regularities about how under-5 mortality is distributed by detailed age in human populations. These models will then be used for evaluating and correcting under-5 mortality information by detailed age in less-developed countries. This global database, and models derived from it, will also allow us to address specific substantive questions about how and why age patterns of under-5 mortality vary by sex, time, and place, with important programmatic implications. This database will be made publicly-available, allowing the research community to easily access high-quality primary information on under-5 mortality by detailed age for their own research needs, and for further validation and replication of the project's results.

Public Health Relevance

Information about how the risk of death varies with age within the 0-5 age range represents critical evidence for guiding health policy. This information indicates ages at which children are particularly vulnerable, helps better evaluate the impact of health interventions, and provides indirect information about underlying causes of death. The present study will make important contributions to our understanding of age patterns of under-5 mortality by bringing together high-quality global information on these age patterns, by providing a new method for estimating them in places where data are deficient, and by improving our knowledge about how and why these age patterns vary by sex, time and place.

SPECIFIC AIMS

The Under-5 Mortality Rate (U5MR) is a key and widely-used indicator of child health, but it conceals important information about how this mortality is distributed by age. For better understanding and monitoring of child health, it is critical to examine how the risk of death varies within the 0-5 age range. This includes age breakdowns beyond the standard cut-off points of 28 days (for neonatal mortality) and 1 year (for infant mortality). In many populations, however, the age pattern of under-5 mortality is not well known. Less-developed countries, in particular, lack the high-quality detailed vital registration information necessary for the analysis of such age patterns. Sample surveys collecting retrospective birth histories do not satisfactorily fill this gap, because they are subject to systematic biases that are particularly consequential for estimating age patterns. This makes the need for high-quality information on age patterns of under-5 mortality even more critical, because regularities in these age patterns can be used as a powerful tool for evaluating and correcting data when sources are deficient.

The goal of this project is to improve our understanding of age patterns of under-5 mortality by gathering the largest database to date on high-quality global mortality information by detailed age (by days, weeks, months, and years of age) from birth until age 5, by sex. This database, which will cover a wide array of historical and contemporary contexts in both more- and less-developed settings, will serve as a basis for generating models summarizing regularities about how under-5 mortality is distributed by detailed age in human populations. These models will then be used for evaluating and correcting under-5 mortality information by detailed age in less-developed countries. This global database, and models derived from it, will also allow us to address specific substantive questions about how and why age patterns of under-5 mortality vary by sex, time, and place. Specifically, we will pursue the following aims:

AIM 1: Develop a baseline model for age patterns of under-5 mortality using high-quality vital registration information from more-developed countries. This will be accomplished by compiling from archival and electronic sources an exhaustive empirical database on under-5 mortality by detailed age for historical and contemporary populations with high-quality vital registration information. This database, which will cover mostly European and North American countries, will then be utilized to develop a flexible two-parameter baseline model, Model A, for use in developed countries or in certain developing countries with similar age patterns.

AIM 2: Update the baseline model with validated prospective data sources from less-developed populations. Additional variation in age patterns of under-5 mortality will be examined by compiling a second database relying on the following prospective sources for various populations in Africa, Asia, and Latin America: (1) Sample Registration Systems; (2) Health and Demographic Surveillance Systems; (3) Cohort Studies; (4) Urban Vital Registration systems. These sources will first be carefully evaluated, and those meeting high quality standards will inform a second model, Model B, for use in populations not well described by Model A.

AIM 3: Use modeled age patterns for the indirect estimation of detailed mortality from birth to age 5 in less-developed countries. Models A & B will be used to evaluate and correct (when appropriate) under-5 mortality information based on sources relevant for many less-developed settings, including: (1) full birth histories from Demographic and Health Surveys; (2) Incomplete Vital Registration systems. This indirect estimation procedure will produce improved estimates, with uncertainty bounds, of under-5 mortality by detailed age in less-developed countries.

AIM 4: Address specific substantive questions about how and why age patterns of under-5 mortality vary by sex, time and place. The rich global database compiled as part of this project, and models derived from it, will allow us to address the following substantive questions: (1) How widespread are exceptions to the usual pattern of regular decline in mortality with age from birth to age 5? (2) Which age groups within the 0-5 age range are most responsive to specific health interventions? (3) How do sex differentials in mortality evolve with age within the 0-5 age range in various contexts?

A by-product of this project will be a publicly-available online database containing all of the primary aggregate data collected as part of Aims 1 & 2 and used for developing models. This global database will allow the research community to easily access high-quality primary information on under-5 mortality by detailed age for their own research needs, and for further validation and replication of the project's results.

RESEARCH PLAN

A. SIGNIFICANCE

The **Under-5 Mortality Rate** (i.e, the probability that a newborn will die before reaching age 5, also denoted U5MR, $_5q_0$, or q(5)) is a key mortality indicator routinely used for tracking progress in the area of child health. As such it was prominently featured among the Millennium Development Goals (MDGs), and it remains an important component of the recently-adopted Sustainable Development Goals (SDGs).[1-4] This indicator, however, conceals important information about how this mortality is distributed by age.[5] For a number of purposes, it is critical to examine how the risk of death varies within the 0-5 age range. This includes not only the standard cut-off points of 28 days (for neonatal mortality, now featured alongside U5MR in the SDGs) and 1 year (for infant mortality), but also detailed information by days, weeks, months and years of age.

Detailed information about the age pattern of Under-5 Mortality (U5M) is important, because it offers critical evidence for **guiding health policy**. **First**, information about how the risk of death varies with age within the 0-5 range indicates ages at which children are particularly vulnerable and where resources may be targeted. A sole focus on U5MR or other standard indicators within the 0-5 range such as the Neonatal Mortality Rate (NMR) or Infant Mortality Rate (IMR) can hide important features of the mortality pattern, including critical ages at which a slowdown or a reversal in the mortality trajectory may be occurring.[6-8] This notion of critical ages for intervention in the mortality literature intersects greatly with the nutritional literature, which emphasizes critical age windows during which child development is particularly amenable to intervention, with a host of long-term health and cognitive consequences.[9-13] **Second**, information on mortality by detailed age within the 0-5 range allows researchers to better evaluate and understand the impact of specific interventions.[14] If child mortality outcomes are measured only in terms of standard, arbitrary age groups such as neonatal, infant and under-5, researchers may miss critical evidence about effects that operate specifically within some non-standard age interval.[15] For example, if some intervention has a specific effect between, say, ages 7 and 15 months, this effect may appear small or irrelevant when focusing on more standard age groups. **Third**, knowledge about detailed age patterns of U5M conveys useful information about a population's epidemiological context in situations where cause-of-death information is unavailable or unreliable.[16-20] Indeed, most causes of death have a clear age pattern which heavily influences how all-cause mortality is distributed by age. At a given level of U5MR, a higher concentration of mortality during the earlier part of the 0-5 age range reflects higher levels of mortality from congenital anomalies and perinatal conditions, since these causes are responsible for most mortality during the early days of life.[19-22] By contrast, infectious diseases have an older age pattern within the 0-5 range, because during the first few months of life a newborn is protected to some extent by the passive immunity inherited from the mother and by breastfeeding.[8,23-26] Diseases such as pneumonia, diarrhea, malaria, and measles all contribute to an older age pattern of U5M, in addition to acting directly on the level of U5MR.[18,20,27] For some of these diseases (especially diarrhea), one important factor influencing age patterns is the length of breastfeeding, because the weaning period represents a particularly vulnerable period with new infectious exposures.[24,28-30] HIV/AIDS is also known to impact the age pattern of U5M.[31,32] The specific effects are complex – they depend on whether babies are infected in utero, intranatally, or postnatally[31] – but they tend to generate an older age pattern of U5M in addition to higher U5MR levels. Besides infectious diseases, injuries also produce an older age pattern of U5M. For example, drowning (which accounts for an important share of under-5 deaths in South Asia) occurs primarily among children above age 1.[33-35] Given these age-specific influences, age patterns of all-cause mortality can act as a powerful proxy for causes of death and help guide policy when this information is lacking. In fact, global estimates of under-5 causes of death are to a large extent based on models which use age patterns of all-cause mortality as inputs.[18,20,36]

In spite of their epidemiological significance, age patterns of U5M are **difficult to establish** in countries that lack reliable vital registration (VR) systems, which includes most Low and Middle Income Countries (LMICs). In these countries, age patterns (including values of NMR and IMR) are derived primarily from birth histories collected as part of sample surveys such as Demographic and Health Surveys (DHS). In addition to sampling errors, data from birth histories are subject to a number of non-sampling errors – including omission of deaths and age misreporting – that are particularly consequential for the estimation of these age patterns.[37,38] For example, most DHS surveys report a disproportionate number of deaths occurring at age 12 months, due to age heaping, making it difficult to distinguish mortality below vs. above the child's first birthday. This means that even the most basic distributional information of U5M (0 vs. 1-4 years) will be

difficult to ascertain in less-developed settings, and in fact, for most LMICs, international agencies do not use raw DHS information on infant mortality in their statistical reports. Omission of deaths in DHS surveys, which are more likely to occur when the child dies soon after birth, represents an even more serious (though less visible) problem.[37,39-41] Not only do these omissions affect how U5M is distributed by age, but they affect the very level of reported U5MR.

In order to address these data deficiencies, researchers sometimes rely on observed regularities about how mortality is distributed by age, using for example **model life tables**. Model life tables, such as the ones developed by Coale and Demeny[42,43] or the United Nations,[44] represent a useful framework because they allow the estimation of arrays of age-specific mortality rates on the basis of only one or two mortality indicators, chosen as entry parameters.[45] When the model correctly represents the population's age pattern of mortality and the entry parameters are accurate, this approach can produce precise results. This approach is in fact commonly used by international organizations to derive IMR on the basis of U5MR. Current usage of model life tables for estimating patterns of U5M, however, is affected by several **important drawbacks**. **First**, existing model life tables only offer 0 vs. 1-4 as an age breakdown for U5M. This is insufficient for most purposes, including for the estimation of NMR or mortality in non-standard age ranges. (One model that contains additional age details is Bourgeois-Pichat's "biometric" model. This model, however, focuses on the first 12 months of age only and has been shown to poorly fit data in a variety of contexts.[46-49]) **Second**, existing model life tables rely on rather old data, with the most recent information dating back to the early 1980's. **Third**, they do not adequately represent the experience of LMICs.[5] In Sub-Saharan Africa, in particular, there is strong indication that U5M follows patterns that are substantially different from those represented in existing models.[5,6,8,25,28,38,50-55] Apart from model life tables, researchers have used approaches that focus specifically on estimating neonatal mortality. This somewhat different line of work relies on the observed relationship between NMR and U5MR across a variety of sources, primarily DHS.[56] This approach, however, does not adequately address data quality issues in the DHS, and is by definition not helpful for the estimation of mortality at ages beyond the first 28 days.

Because of these deficiencies in data and methods, tracking global progress in child health has not been as successful as expected at this stage of the data revolution. Another consequence of these deficiencies is that a number of important **substantive questions** in the area of child health remain poorly answered in the literature. Three questions stand out. The **first** question pertains directly to the shape of the mortality curve between ages 0 and 5. One assumption often made in this literature is that the risk of death within the 0-5 range is highest during the first 24 hours and then declines continuously until age 5. This pattern, observed in many settings, has been interpreted using evolutionary models[57,58]. However some important exceptions have been pointed out in the literature, including in some West African populations where the risk of death, after an initial decline, makes a sharp reversal around age 6 months.[7,59-61] This reversal appears to be related to a combination of factors, including inadequate weaning foods.[8,26,28] It remains unclear, however, how widespread these exceptions are, due to a lack of reliable data. Additional information about the geographical distribution of this pattern, and the ages at reversal, would have critical theoretical and programmatic implications. The **second** question focuses on which age ranges within the 0-5 range, including non-standard age groups, are most responsive to specific large-scale immunization programs and other health interventions. While the literature has mostly focused on standard age ranges, more precise knowledge about specific age windows at which interventions are most effective has important methodological, epidemiological and programmatic implications.[14,15] The **third** question is in reference to sex differentials in mortality. While there is a wide literature on sex differentials in infant and child mortality in various contexts, few researchers have examined these differentials using finer age information. More precise evidence about how sex differentials evolve within the 0-5 age range, starting with the first 24 hours when genetic and biological factors are the most salient, would greatly inform debates about underlying processes.[21,62,63]

This project proposes to fill these important gaps by gathering the most comprehensive **database** to date on high-quality worldwide mortality information by detailed age (days, weeks, months, trimesters and years as explained in Section C.1) between 0 and 5 years. The database will include information from both historical and contemporary periods, and from both more- and less-developed contexts, thus offering a large spectrum of epidemiological situations for examining regularities in how U5M is distributed by age in human populations. This database, which will be made publicly available, will be used in two different ways as part of this project. First, it will be used to **develop models** for the indirect estimation of mortality by detailed age

from birth to age 5 in situations where data are of questionable quality, including birth histories from DHS and incomplete vital registration data. Second, it will be used for generating new knowledge about the **substantive questions** discussed above. The focus of this project is on all-cause mortality. Obtaining reliable, comparable information on causes of deaths for such detailed age groups in so many different contexts would be a monumental task and is beyond the scope of this project. Nonetheless, observed patterns will be very much discussed with reference to causes of death in this project, due to the important role played by causes of death in determining patterns of all-cause mortality.

B. INNOVATION

• This project is innovative on many different levels. **First**, it will generate a **database** on high-quality empirical patterns of U5M that will be unprecedented in terms of amount of age information and breadth of temporal and geographical coverage. By contrast, the Human Mortality Database (HMD) does not provide as much detailed age information (single years at most), and does not cover LMICs. This new database will be made publicly available over the course of the project.

• **Second**, this project proposes a new **modeling approach** for summarizing and predicting age patterns of U5M. This model is innovative both in terms of its empirical basis and its specifications. Existing models in this area are hindered by the limited geographical scope of their empirical basis, or by their inclusion of data sources of questionable quality. Also, existing models either focus on broad age groups, or on only one detailed age group like neonatal mortality. In this project, we believe that much can be learned by taking all age groups within the 0-5 range into account at once, because patterns at one age tend to be correlated with patterns at other ages in a given population. Also, many data errors appear clearly only once the entire age distribution of mortality is taken into account.

• **Third**, this model will provide a new, powerful strategy for **evaluating and correcting** U5M information in LMICs. This new approach will be useful for evaluating information from DHS, a data source that is often taken for granted in this literature in spite of its weaknesses. This approach will also be useful for evaluating incomplete VR information in certain situations. Incomplete VR information is often dismissed entirely as a data source, even though it can contain valuable information for U5M estimation. Finding new solutions for evaluating and correcting incomplete VR information is an important goal in itself, because of the new international impetus to strengthen civil registration and vital statistics in LMICs, and because countries often find it more empowering when their mortality estimates are based on their own data, even after adjustments, rather than on international survey programs such as the DHS.[64]

• **Fourth**, our model and evaluation procedure will produce new, **improved estimates** of U5M patterns in many LMICs. We anticipate that these improved estimates will generate many surprises, with some countries showing significantly different levels and trends in U5M than previously believed.

• **Fifth**, the new data and methods generated by this project will provide **new answers** to important substantive debates in the literature, with the potential to seriously challenge existing knowledge in this area. We also expect that other researchers will advance knowledge on these and other debates by using the publicly-available database that will be generated as part of this project.

C. APPROACH

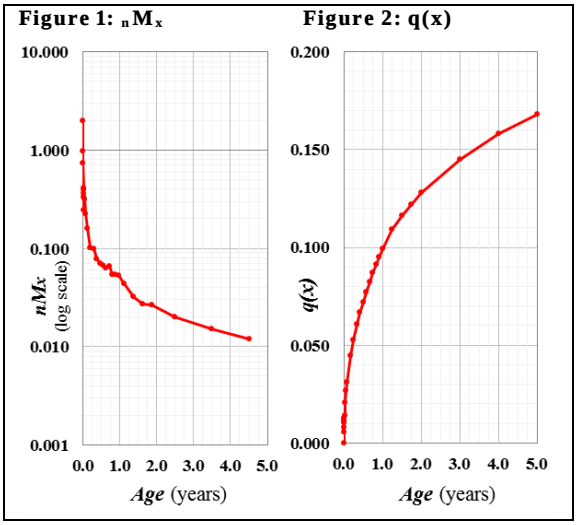
C.1. Generalities

• **Live births.** Proper measurement of U5M starts with a clear and consistent definition of what constitutes a live birth. In this project, we will rely on WHO's definition, which takes any sign of life following delivery, regardless of gestational age, as an indication that a birth is a live birth. This means that any data source that does not track live births according to that definition will need to be treated as potentially problematic. This includes vital registration systems in countries that have followed different rules for reporting live births (e.g., former Soviet republics),[65,66] or retrospective surveys in which respondents may not be able to properly make the distinction between live births and stillbirths.

• **Detailed age groups.** The age groups that we will most consistently use in this project have the following beginning and end points (in exact age): 0 day (birth), 1d, 2d, 3d, 4d, 5d, 6d, 1 week, 2w, 3w, 4w, 1 month, 2m, 3m, 4m, 5m, 6m, 7m, 8m, 9m, 10m, 11m, 12m, 15m, 18m, 21m, 2 years, 3y, 4y, 5y. These are the cut-off points that we most consistently find in the sources we identified. This means that we will work with distributions of deaths in age groups defined by any two consecutive cut-off points, starting with deaths occurring between birth and the end of the first day (first 24 hours), and ending with deaths occurring between a child's 4th and 5th

birthday. The width of each age interval, *n*, will vary from n=1 day, 1 week, 1 month, 1 trimester to n=1 year starting with the 2nd birthday.

• **Mortality measures.** This project focuses on two different mortality measures for analyzing how U5M is distributed by age. Our first measure is ${}_nM_x$, the mortality rate between age x and x+n. This rate is defined as the number of deaths in the interval x to x+n per person-year lived in that interval. ${}_nM_x$ is an approximation of the force of mortality (μ(x)) and shows how the risk of death varies with age. Most consistently in this project, we will calculate ${}_nM_x$ by combining distributions of deaths by detailed age with distributions of the corresponding population by single year of age, using the assumption that the population is uniformly distributed within each single-year age group. Our second measure is q(x), the cumulative probability that a newborn will die before reaching exact age x. q(x) trajectories from birth to age 5 show how a certain value of U5MR (=q(5y)) is progressively reached in a given population, passing through age-specific values including the early neonatal mortality rate (q(1w)), the NMR (=q(4w)) and the IMR (=q(1y)). Most consistently in this project, we will calculate q(x) on the basis of ${}_nM_x$ with the assumption that the force of mortality is constant within the age interval x to x+n (a non-consequential assumption considering the very small width of our age intervals, particularly for the younger ages at which mortality is decreasing rapidly). As an illustration, Figures 1 & 2 present ${}_nM_x$ and q(x) trajectories for Swedish females in 1891. The ${}_nM_x$ trajectory shows that in this population the risk of death is highest during the very first day of life, followed by a monotonic decrease until the 5th birthday. The q(x) trajectory shows how the corresponding cumulative probability of dying increases with age, with a NMR of .031, an IMR of .099 and a U5MR of .168. Trajectories of μ(x) (approximated with ${}_nM_x$) and q(x) are related to one another mathematically with the following equation: $\mu(x) = -d[\ln(1-q(x))]/dx$. In this project, we will switch back and forth between our two measures, depending on modeling needs and features of the age pattern we want to highlight.



• **Period vs. Cohort measures.** We will focus on period (i.e., synthetic cohort) measures of ${}_nM_x$ and q(x). For most of the data sources we use, we don't have the needed information to calculate measures for real cohorts. Important exceptions are Health and Demographic Surveillance Systems (HDSS, see Section C.3.ii) and the Cohort Studies (see Section C.3.iii). Given that the focus of this proposal is on a segment of the life cycle that spans a rather limited time frame (5 years), we do not expect age patterns to vary significantly by period vs. cohort, but will examine this contrast in HDSS for which both period and cohort measures can be calculated. Most consistently in this project, ${}_nM_x$ and q(x) will be calculated over time periods with a width of 1 year, but wider periods will be used when dealing with samples.

• **Sex-specific vs. both sexes combined.** Even though indicators of child mortality are most often presented for both sexes combined in the global health literature, our project will pay specific attention to patterns by sex. Data sources for which the raw information is not provided by sex will not be included, and all our models will come in 3 versions: females, males, and both sexes combined.

C.2. Baseline model for age patterns of under-5 mortality in more-developed countries (Aim 1)

This first aim focuses on historical and contemporary patterns in more-developed countries. The reason for this focus is the existence in these countries of long time series of high-quality, national-level information on deaths by detailed age, gathered by vital registration systems. This gives us a solid basis for developing a first set of models, which will be updated and expanded at a later stage with information from LMICs (Aim 2).

• **Development of a database for high-quality VR countries.** The database developed as part of Aim 1 will build on the existing Human Mortality Database (HMD).[67] The HMD provides historical and contemporary mortality information for 38 countries (located primarily in Europe, North America, and Oceania) where death registration and census data are virtually complete, as determined by stringent quality criteria. Due to its focus on mortality over the entire life course, this database contains fewer age details than we need below age 5 (single years of age at most). Nonetheless the HMD is useful for our purpose, because it includes a large portion of the primary data we need for our own database, including Jan 1 population estimates by single-year

age groups, annual deaths by single-year age groups above the 2nd birthday, and annual births. This means that the only additional data that we will need to collect are distributions of deaths by detailed age below the 2nd birthday. These distributions of deaths are readily available electronically for all these countries since 1970 through a database gathered by the United Nations which we have already obtained. For years prior to 1970, these needed tables are available in archival sources and need to be digitized. These sources, which mostly consist of statistical yearbooks, are listed in the HMD and available in various libraries in the US. This archival data collection, a portion of which we have already compiled as preliminary work for this proposal, will be completed during the course of the project. The final database will include mortality information from more-developed countries for years as early as 1800, all the way to the present. This represents a period of rapid mortality change, from situations of high U5MR (>200 p.1000) in which populations have little control over infectious diseases, to situations of low U5MR (<5 p.1000) where avoidable deaths have been largely eliminated and where almost all remaining deaths are due to congenital anomalies. Using the collected information, we will calculate for each country-year ${}_nM_x$ and q(x) values by detailed age and sex, following the procedure explained in C.1. The final product of this portion of the project will be a database for mortality in the 0-5 age range with a similar geographical and temporal coverage as the HMD, but with age details not available in the original HMD.

• **Modeling approach.** Our goal is to develop a model for the indirect estimation of mortality by detailed age from birth to age 5. Building on an approach developed by Wilmoth et al.[68] for estimating full life tables, we propose the following 2-parameter log-quadratic model:

$$\mathbf{ln[q(x)] = a_x + b_x\, ln(q(5y)) + c_x\, ln(q(5y))^2 + v_x\, k} \qquad \mathbf{(1)}$$

This model, which uses q(5y) as main predictor, is a system of equations, with one equation for each detailed age x (1d, 2d, ..., 4y). q(5y) determines the level of mortality, while k affects the shape of the age pattern at a given level of q(5y). This model is justified by the observation of log-quadratic relationships between q(5y) and q(x) at various ages x in the preliminary database we gathered, and the fact that when k=0, prediction errors tend to occur in a similar fashion across age groups for the same population. Co-variation across ages of such errors is captured by v_xk, where v_x depicts the age pattern of typical deviations from expected q(x) when k=0.

This model will be fitted in two steps. First, a_x, b_x and c_x will be estimated simultaneously for each age x using least-squares methods. Second, v_x will be estimated as the first term of a Singular Value Decomposition (SVD), computed from the matrix of regression residuals. Once a_x, b_x, c_x and v_x will be estimated, the model can be used for demographic estimation, predicting the entire q(x) trajectory for a given population on the basis of one or two entry parameters. For example, one may use q(5y) as the first entry parameter, and then solve for the value of k that matches the desired second entry parameter (say, q(5m)). Expected values of q(x) can then be estimated from q(5y) and the matched value of k. Finally, using the set of predicted q(x) values, mortality indicators between any two available ages within the 0-5 range can be estimated, including NMR, IMR, or mortality between, say, 7 and 15 months. The model can also be used with any pair of mortality indicators within the 0-5 age range as predictors.

• **Preliminary results**. As preliminary work for this proposal, we estimated a_x, b_x, c_x and v_x in Equation (1) using the preliminary database that we already collected (which includes the following country-years: Belgium 1841-, Denmark 1890-, England & Wales 1908-, Finland 1881-, France 1899-, Germany 1950-, Japan 1950-, Netherlands 1848-, Norway 1875-, Portugal 1929-, Sweden 1891-, and US 1933-). Figure 3 shows results from the female model, illustrating the effect of changing q(5) on q(x) trajectories when k=0. As expected, the portion of U5MR that occurs below age 1 increases as U5MR decreases. Figure 4 shows the effect of changing k when q(5)=.100. At a given level of U5MR, negative values of k represent an earlier age pattern of mortality (with a greater portion of mortality occurring at earlier ages), while positive values represent a later age pattern.

Figure 3: level effect

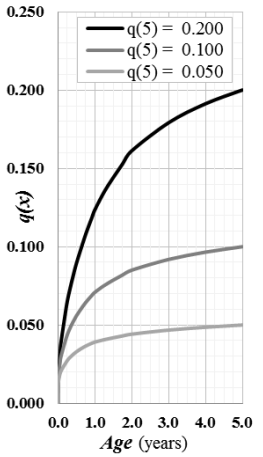
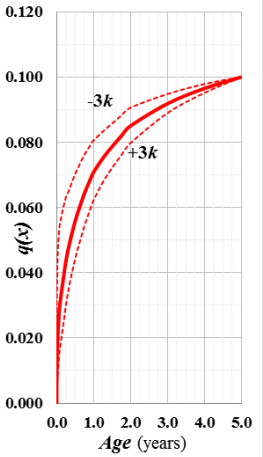


Figure 4: shape effect



• **Generating Model A.** The final goal of Aim 1 is to generate a version of our model that will be based on the complete database collected as part of this aim. A thorough sensitivity analysis will be conducted to determine which model specification and estimation procedure produce the best fit. We will also

determine which combination of entry parameters best reproduces the data. The final model will offer a set of coefficients a_x, b_x, c_x and v_x that best represent regularities in the age pattern of U5M in more-developed countries, for each sex and for both sexes combined. Although Equation (1) is highly promising, our attachment to it is initial at this stage, and we plan to experiment with other forms once the database is complete.

• **Limitations.** While the mortality data for HMD countries represent a gold standard in terms of quality, coverage of under-5 deaths in some country-years may still be questionable, especially during earlier years and at very early ages. This problem will be alleviated by the exclusion of country-years identified as potentially problematic in the HMD documentation (including cases in which proper distinction between live births and stillbirths may not be fully warranted, like France before 1899)[69]. Additional exclusion of country-years may be performed if the new death information by detailed age indicates errors that the HMD researchers did not see when they examined information by broader age group.

C.3. Modeling age patterns of under-5 mortality in less-developed settings (Aim 2)

Model A will be useful in the sense that it will be based on high-quality information over extended periods of time during which populations have seen huge changes in epidemiological environments. A drawback, however, is that it won't involve information from LMICs. In order to increase the model's external validity, we propose in Aim 2 to update it with high-quality information from a variety of less-developed contexts. This information will be based on various data sources from less-developed populations scattered around the world, covering a variety of epidemiological environments. These data sources include: (i) Sample Registration Systems (SRS); (ii) Health and Demographic Surveillance Systems (HDSS); (iii) Cohort Studies (CS); and (iv) Urban Vital Registration (UVR) systems. Within these sources, we will only examine data sets for which we have access to raw, unadjusted information. This will ensure that all mortality calculations are made following a standard set of procedures. One common characteristic of these four sources is that they are all prospective, a criteria that is justified by our need for high completeness of vital events and precise age information. We are thus purposely excluding retrospective data sources, including DHS, for reasons explained earlier. (DHS data will be treated separately in Aim 3.) In spite of their potential for producing high-quality information, the sources examined in this aim have not been as systematically evaluated as in the case of the HMD, and we anticipate that their quality will be more variable. Therefore, a large portion of Aim 2 will consist of a thorough data quality evaluation. Only data sets that meet objective data quality criteria will be used to inform models.

Except for SRS, sources in Aim 2 are not representative of patterns at the national level. We do not see this as an important drawback, because the purpose of examining these data sources is not to produce national-level mortality estimates. Rather, the purpose is to study how mortality is distributed over ages given a certain level of mortality in a given epidemiological context. A given pattern for a well-defined sub-population, even if not representative, will uncover certain features of age patterns that reflect underlying epidemiological processes and thus need to be taken into account in models. For example, when a sub-national sample presents an increase in age-specific mortality around age 6 months, it shows that monotonic declines in age-specific mortality cannot be a general, fixed property of models.

Our data quality assessment will be tailored to some extent to the type of source, since different sources have different data collection protocols with specific weaknesses. However, some internal consistency checks will be common to all sources. First, we will discard site-years that have a significant pattern of age heaping. This is justified by the fact that mortality by detailed age is sensitive to age heaping, and that age heaping is often correlated with broader data quality issues, including underreporting. We believe this is a rather stringent but necessary criteria given the goals of the project. Second, we will discard site-years for which age-specific mortality presents a pattern of systematic increase by age during the first seven days or weeks of life. While some increases in mortality have been documented later in the 0-5 range, such as around age 6 months, there is neither empirical nor theoretical evidence indicating that the true risk of death can be higher, say, during the 3rd day than during the first day, or that it can be higher during the 3rd week than during the first week. Such observations will be taken as a strong indication that the data are erroneous and will not be included in the final database. Unlike the exhaustive population information examined in Aim 1, most of the data sources examined in Aim 2 will be affected by sampling error. This sampling error will be taken into account before making a decision about whether a given data set meets the above data quality criteria. Only those data sets for which suspicious patterns cannot be explained by sampling error will be excluded. In the next sections, we describe our four data sources and our approach for analyzing them.

i. Sample Registration Systems (SRS). Sample Registration Systems are large data collection operations conducted by governments to provide high-quality continuous demographic information in situations where classic VR systems are incomplete or inexistent. SRS are conducted over sampled geographic areas chosen in a way to ensure that the resulting information will be representative at the national level (and sometimes also at the sub-national level). For this project, we will analyze detailed SRS information for the world's two largest countries, China and India, which alone cover more than a third of the world's population.

• **China's Maternal and Child Mortality Surveillance System (MCMSS).** MCMSS is a population-based surveillance system with a representative sample of districts/counties scattered across China (excluding Hong Kong, Macau and Taiwan).[70,71] This system was set up in 1996 with 123 districts, and then expanded to 336 districts since 2007. In 2014 alone, the MCMSS tracked a total of 431,610 live births and 4,374 under-5 deaths. The MCMSS is the primary source of U5M information in China. It is designed to produce national-level estimates as well as estimates for 6 broad regional strata. The data collection procedures include thorough quality control at all levels of the surveillance system, ensuring in principle a virtually complete coverage of births and under-5 deaths in the sampled areas. In the context of this project, we have established a collaboration with MCMSS staff, who will provide us with the detailed raw information we need (annual counts of live births, under-5 deaths by detailed age, and population by single years of age, by sex and regional strata, with sampling weights). This information will allow us to examine for the first time patterns of ${}_nM_x$ and q(x) by detailed age in China, for each year since 1996. These mortality indicators will be calculated for each sex, at the national level and for each of the 6 broad regional strata. The MCMSS data will be evaluated using the internal consistency checks discussed above.

• **India's Sample Registration System (SRS).** Like China's MCMSS, India's SRS involves sample units that are randomly selected to be representative at both the national and subnational (state) level.[72] Created in 1971, the SRS now includes 8,861 sample units scattered in all 28 states and 7 union territories of India. An average of 150 households are drawn within each sample unit. These selected households are continuously monitored for vital events using a thorough dual-record system. This system presumably ensures a high-level of completeness of vital events. In 2013, the SRS tracked a total of about 150,000 live births. For this project, we will work in collaboration with SRS staff to produce for the first time national-level estimates of U5M by detailed age in India. Specifically, values of ${}_nM_x$ and q(x) by sex and detailed age will be calculated for each year from 2014 onwards. Annual data will be provided progressively throughout the project, with 2020 as the final year. Like China's MCMSS, these data will be evaluated using internal consistency checks.

ii. Health and Demographic Surveillance Systems (HDSS). HDSS sites are circumscribed, sub-national geographic areas within which complete longitudinal monitoring of the population is performed for research purposes.[73] Demographic monitoring within an HDSS site typically starts with a full census of the site's population. Households in the area are then visited at regular intervals, with detailed tracking of births, deaths, in-migrations and out-migrations that occurred since the last visit. For births and deaths that occurred between visits, questions are asked about corresponding dates. Frequent visits mean that there is little room for births and deaths to be unreported or be subject to errors in dates. HDSS sites have generated a large amount of useful information on many subjects, including health, mortality, fertility, migration, and family structure. HDSS-based mortality information has also been used for mortality modeling, including the development of model life tables for Sub-Saharan Africa.[74] HDSS sites represent a rich data source for our purpose. Some HDSS-based studies have in fact detected peculiarities in the age pattern of U5M in specific areas, motivating us to include HDSS information in this project.[5] The quality of the U5M information in HDSS sites, however, may vary across sites, depending on site-specific data collection practices. Certain sites exhibit U5M rates that are implausibly low. The implication for our project is that HDSS mortality information should not be used indiscriminately. Rather, we propose to conduct a thorough evaluation of the quality of HDSS data for estimating U5M estimation by detailed age. This will be conducted in two stages. First, we will take advantage of micro-data available in a data repository for 14 HDSS. Second, we will conduct in-depth analyses from another 8 study sites with which we have an agreement to access additional micro-data.

• **General analysis of HDSS micro-data available on data repository.** Micro datasets from a selection of HDSS sites are publicly available in a data repository called iSHARE and managed by INDEPTH, an umbrella organization for HDSS sites throughout the world. These datasets, which are listed in the table below, contain the necessary data for calculating the needed exposure terms, age distributions of deaths, and

Study name	Years available	Under-5 population (2012)	U5MR (2012)
Asia			
Vadu (India)	2009-2012	13,164	19
Africa			
Agincourt (South Africa)	1992-2012	11,077	36
Dabat (Ethiopia)	2008-2012	6,136	79
Dikgale (South Africa)	1995-2012	4,007	16
Gilgel Gibe (Ethiopia)	2005-2012	9,673	77
Iganga/Mayuge (Uganda)	2005-2012	12,628	70
Karonga (Malawi)	2003-2015	6,486	73
Kersa (Ethiopia)	2007-2012	8,524	105
Kilte Awlaelo (Ethiopia)	2009-2012	7,060	33
Mbita (Kenya)	2009-2011	7,504	45
Nairobi (Kenya)	2002-2012	9,615	66
Ouagadougou (B. Faso)	2009-2012	13,176	36
Rufiji (Tanzania)	1998-2012	13,816	63
Taabo (Côte d'Ivoire)	2009-2012	6,034	89

corresponding ${}_nM_x$ and q(x) values by detailed age within the 0-5 age range. These indicators will be calculated over five-year periods or less, depending on the available years. We will then perform the internal consistency checks discussed above. Observed irregularities will be examined in light of objective site-specific data collection criteria, including frequency of household visits and the existence of detailed probes for early deaths in data collection instruments. We will pay particular attention to sites in which women who were not pregnant at the last visit are filtered out of questions on birth outcomes, a situation which leads to underestimation of early deaths if visits occur only annually. We anticipate that this data evaluation will raise a number of concerns about data quality. These concerns will be further evaluated in light of the results of a more detailed analysis of select sites, as explained in the next section.

• **Detailed analysis of select sites.** We will perform a detailed analysis of 8 sites (listed in the table below) that have agreed to furnish more detailed data. These data will facilitate the estimation of U5M patterns and clarify possible sources of bias. These sites are selected for several reasons. First, they offer long periods of demographic monitoring, allowing the analysis of time changes in the age pattern of U5M, sometimes during periods of rapid mortality decline. Second, these sites have been used for clinical trials and other interventions, allowing us to study how interventions impact age patterns of U5M (see Aim 4). Third, these sites represent a range of epidemiological environments, with U5MR varying from 31 to 104 p.1000. Fourth, these sites follow enhanced data collection protocols, which include supplementary data collection on the identity of the respondent who reports on behalf of the household in each census round, verbal autopsy data on dates of birth and death, pregnancy status information (as reported during census visits or via individual-level record linkage with health facilities offering antenatal care (ANC) services in the HDSS area), retrospective birth histories, and information retrieved from child health cards. These features will allow us to study the impact of enhanced vs. standard data collection protocols on patterns of errors, information which will then be useful for further assessing the quality of the mortality information for the iSHARE sites. For each of these sites, we will calculate ${}_nM_x$ and q(x) for five-year periods (as well as five-year cohorts), by sex. After performing data quality checks similar to those performed for the iSHARE sites, we will examine the impact of record linkage on estimated mortality rates, by comparing mortality levels and age patterns with and without inclusion of information generated by the enhanced data collection features. This will allow us to better understand patterns of under-reporting in sites that do not perform record linkage. We will also examine bias arising from missed pregnancies occurring when the frequency of visits is low (annual or less frequent). This will be done by removing from the original data the births and deaths that would have been missed had the sites used standard rather than enhanced instruments. Results from these detailed analyses will allow us to better understand irregularities in the iSHARE data sets and further motivate decisions about inclusion vs. exclusion of individual data sets. The final product of these general and detailed analyses of a total of 22 HDSS sites will be a series of high-quality U5M rates by detailed age for a variety of sites in LMICs, with a special emphasis on Sub-Saharan Africa where the true age patterns are the least well known.

Study name	Years of systematic follow-up	Under-5 population (2012)	U5MR (2012)
Asia			
Matlab (Bangladesh)	1966	24,215	33
Africa			
African Centre (South Africa)	2000	9,250	38
Bandafassi (Senegal)	1970	2,300	104
Bandim (Guinea-Bissau)	1978	37,402	88
Basse/Faraffeni (Gambia)	1981	40,027/11,481	40/47
Magu (Tanzania)	1994	5,767	46
Mlomp (Senegal)	1985	896	31
Niakhar (Senegal)	1962	7,960	71

iii. Cohort Studies (CS). Cohort Studies of children below age 5 offer a rich, yet underused source of information for mortality estimation. The main purpose of most CS is not to estimate levels of child mortality, but to measure the effect of interventions or other factors on various child outcomes, sometimes using experimental designs. Because of these specific objectives, researchers conducting CS typically invest a large amount of resources in tracking subjects over time, through household visits that are far more frequent than in

any of the other data sources examined in this project. These studies thus present a level of thoroughness in their mortality follow-up that is unparalleled in less-developed settings.[75] The flipside of such intensive research designs is that samples are sometimes not very large and not representative of broader geographical areas. Also, these studies often unfold over a somewhat limited time frame, preventing the analysis of time dynamics. For the purpose of this project, however, the advantages of CS far outweigh their limitations. As said earlier, representativeness at the national level does not represent an important drawback for the purpose of this project. As for sample size, we will focus here only on those studies with relatively large sample sizes and/or high mortality levels. Another advantage is that CS with experimental designs can be used in Aim 4 for understanding how age patterns respond to specific interventions. We will focus in this project on 8 different cohort studies, chosen for their sample sizes as well as for the duration and thoroughness of their mortality follow-up. Also, these studies offer a range of epidemiological environments, with IMR varying from 14 to 166 p.1000. For each of these studies, we will calculate cohort-specific ${}_nM_x$ and q(x) functions, by sex. Like in the case of the other datasets, we will conduct internal consistency checks, taking sampling errors and specificities of the field operations into account.

Study name	Setting	Population represented	Number of births followed	IMR	Systematic follow-up period
Asia					
Nepal (1999)	Rural Sarlahi, Nepal	Population-based recruitment of all pregnant women in study area	4,130	67	>5 years
Philippines (1983)	Urban Cebu, Philippines	Population-based, random cluster sample of census	3,080	36	>5 years
Africa					
Burkina Faso (2004)	Hounde, Burkina Faso	Prospective, community based cohort	1,373	67	1 year
Kenya (1995)	Rural Asembo, Kenya	Community-based recruitment by census/TBAs	1,828	166	>5 years
Zimbabwe (1998)	Urban Harare, Zimbabwe	Facility-based recruitment, 14 maternity clinics and hospitals	14,110	93	2 years
Brazil (1982)	Urban Pelotas city, Brazil	Population-based, all births in area hospitals (100% facility delivery)	5,914	28	>5 years
Brazil (1993)	Urban Pelotas city, Brazil	Same as above	5,279	14	>5 years
Brazil (2004)	Urban Pelotas city, Brazil	Same as above	4,287	17	>5 years

iv. Urban Vital Registration (UVR) systems. While vital registration systems in less-developed countries are often incomplete at the national level, urban areas within these countries sometimes have levels of completeness that are in par with those observed among more-developed countries. Although not representative at the national level, these high-quality UVR data provide a rare opportunity to examine mortality patterns in settings that usually do not offer such exhaustive information. In this project, we propose to examine **Antananarivo** (Madagascar), an African city where such high-quality information is available. Antananarivo has a vital registration system that dates back to late 19th century.[76] Completeness has been evaluated as virtually complete for years as early as 1960. This high completeness appears to be related to the fact that cemeteries in the city are fenced and guarded, and that death certificates are needed to obtain burial permits. Death registers in Antananarivo have been digitized for the period 1976-2012, producing a database of about 298,000 individual death records (with detailed ages at death). In this project, we will first update this database for the period 2013-15 in collaboration with Antananarivo's statistical office, and then combine the entire series with available population information for the same period. This will give us about 40 years of detailed raw mortality data for the capital of one the world's poorest country. Although we anticipate high-quality information from this source, we will nonetheless take advantage of the raw information and run the battery of internal consistency checks described earlier. The resulting, final series of annual mortality indicators may be shorter than the original one.

Our examination of UVR systems will also include information available in the **UN database** mentioned in Section C.2. This database includes detailed death distributions by urban/rural residence for many countries. Combining urban deaths with urban population information, we will study patterns of urban mortality in the following countries which have been identified as having a high level of UVR completeness since at least the 1970s: Colombia, Costa Rica, Mexico, Sri Lanka, and Thailand.[44] Specifically, we will calculate urban patterns of ${}_nM_x$ and q(x) by detailed age and sex for each year since 1970, the start year for the UN database. Like the other sources, these patterns will be subject to internal consistency checks.

v. Informing model age patterns with validated information from less-developed countries. Ascertained mortality information from the above sources will then be used to inform mortality models. This

will be done as follows. For each validated site-year, we will examine the extent to which Model A fits observed q(x) functions (using the best pair of entry parameters, as determined in Aim 1). Patterns of prediction errors in q(x) will be analyzed, and measures of goodness of fit will be calculated. Data sets for which prediction errors do not indicate systematic patterns of deviation will be integrated into the database we produced in Aim 1. Model A will then be updated with this information, with regression approaches that take into account the sampling error of this additional information. For ease of exposition in this proposal, we refer to the final, augmented version of Model A as Model A'.

Site-years for which prediction errors indicate a significantly distinct age patterns will be treated separately. They will be combined in a separate database which will form the basis for calculating a distinct set of coefficients in Equation (1). We anticipate that this separate database will consist primarily of some Sub-Saharan populations for which the force of mortality presents an increase somewhere in the age range 0-5. We also anticipate that we will be able to summarize such patterns with one single model, which we call Model B. However, if a single model does not produce an acceptable fit due to significantly different age patterns across populations, we will further stratify the database and produce separate models as needed. Given the flexibility of our model and the continuous nature of the shape parameter k, we do not anticipate this to be necessary. By the end of Aim 2 we anticipate the following two models: (1) Model A', an updated version of Model A for use in more developed countries and some less-developed settings; (2) Model B, for use in some less-developed populations with an unusual pattern of U5M. Here also, Models A' and B will undergo a thorough sensitivity analysis for determining which estimation procedure and entry parameters best reproduce the data. We anticipate that these two models will correctly describe variation in age patterns of U5M in the vast majority of populations worldwide.

C.4. Using models for indirect estimation (Aim 3)

Models A' and B will give us a new tool for evaluating and, when necessary, correcting U5M information by detailed age in sources affected by non-sampling errors. In this project, two such sources will be examined: (1) Demographic and Health Surveys; and (2) Incomplete Vital Registration systems. The underlying logic of our approach is the following. Uncorrected mortality patterns in these sources will be compared with the regularities represented by Models A' and B. Departures from these regularities that are statistically significant will be taken as indication that non-sampling errors are operating, in which case the models will be used as a basis for performing corrections. The final outcome of this aim will be a new set of mortality estimates by detailed age for each country-year covered by DHS or IVR (with uncertainty bounds), as well as general guidelines for how to evaluate these sources for the estimation of U5M in the future.

• **Demographic and Health Surveys (DHS).** As mentioned in Section A, ${}_nM_x$ or q(x) trajectories calculated on the basis of DHS surveys are affected by 2 major sources of non-sampling errors: (1) omission of deceased children; (2) age heaping in ages at death. Our proposed indirect estimation procedure takes advantage of the fact that these errors do not affect all ages equally within the 0-5 range. Omission of deceased children disproportionately affects mortality rates during the early portion of this age range. Age heaping primarily affects ages within a relatively narrow age range around round-digit ages. As a result, these errors generate distortions in age patterns that are clearly identifiable (including, for example, discontinuities in q(x) trajectories). Interestingly, some mortality indicators within the 0-5 range are more robust to these types of non-sampling errors than others. For example, omission of children who deceased below, say, age 2 months will not affect the observed probability that a child aged 2 months will die by their 5th birthday. Similarly, if some deaths occurring between, say, ages 7 and 15 months are misreported as deaths occurring at age 12 months due to age heaping, this will not affect the observed probability that a 7-month old child will die prior to reaching age 15 months. The age groups at which observed mortality is more robust can thus be used as entry points in Models A' and B to produce corrected estimates of detailed mortality within the 0-5 range.

Specifically, we propose the following strategy, which we will apply to every single available DHS (227 surveys covering 78 LMICs throughout the world). First, we will calculate observed ${}_nM_x$ and q(x), by sex, for 10-year periods in order to minimize sampling errors. We will first detect obvious errors in the data, such as peaks in mortality around heaping ages, as well as situations in which mortality increases with age during the first few days and weeks. We will then compare for each country observed vs expected q(x) trajectories using Model A' or Model B, successively using two different pairs of entry parameters: (1) q(5y) and the best second parameter determined in Aim 1; (2) a pair of entry parameters that are robust to errors of omission and age

heaping. In situations where only age misreporting is operating, we expect that the two options for entry parameters will both produce a good overall fit in either Model A' or B, with deviations occurring mainly around heaping ages. The correction will then be done using either pair of entry parameters. In situations where there is significant underreporting of mortality at early ages, the predicted level of q(5y) when using the second pair of entry parameters will be significantly higher. In this case, the data will be adjusted for both underreporting and age heaping, with this second pair of mortality indicators as entry parameters. We will choose Model A' or Model B for making the correction, depending which model produces the smallest prediction errors in the age range covered by the second set of entry parameters. Uncertainty bounds for predicted q(x) values will be calculated using bootstrapping approaches, taking into account sampling error in both the entry parameters and the model's coefficients.

• **Incomplete Vital Registration (IVR) systems.** U5M patterns based on IVR information are affected by sources of errors that have many similarities with those affecting DHS information. Deaths that occur soon after birth are more likely to be unreported in both birth and death registration systems than later deaths. Ages at death are likely to experience some amount of age heaping, especially if the age at death is declared rather than calculated on the basis of dates. Countries that have less-than-complete but nonetheless functioning VR systems are likely to exhibit irregularities in reported U5M patterns that are clearly identifiable and limited to specific portions of the 0-5 age range, as demonstrated for the case of Kyrgyzstan.[66] For such countries, the same procedure as the one described above for DHS data can be applied. In this project, we propose to evaluate and correct VR information for a set of LMICs with a medium Vital Statistics Performance Index (VSPI) score.[77] Among these countries, we will focus on those that also have at least one DHS survey available, so that IVR and DHS information, each adjusted, may be compared. The needed data for these countries, which are primarily located in Eastern Europe, Central Asia, Western Asia, South-East Asia and Latin America, are already available in the UN database described in C.2.

• **Limitations.** This approach to evaluating and correcting DHS and IVR data rests on three main assumptions. The first assumption is that omission of deaths occurs only outside the age intervals used as entry points in the model. The second assumption is that age misreporting of deaths does not cross the age boundaries of the age intervals used as entry points in the model. These first two assumptions will be evaluated directly using the detailed HDSS information collected in Aim 2. Indeed, most of these HDSS have collected retrospective full birth histories in addition to the usual prospective information. Comparing these retrospective vs. prospective reports at the individual level will allow us to better understand patterns of omission and age misreporting, and examine which entry parameters for our models are most robust to common errors. The third assumption is that the age pattern of U5M in the studied population will be well described by Models A' or B. While this may not be always the case, we believe that the breadth of the empirical inquiry in Aims 1 & 2 will provide us with models that cover the vast majority of epidemiological environments. Also, whatever the remaining gaps, these models will represent a huge improvement relative to existing ones, including Coale & Demeny and UN model life tables.

C.5. Addressing specific substantive questions (Aim 4)

The rich global database collected for this project, and models derived from it, will allow us to address a number of important substantive questions in the area of child health. We will first conduct a general examination of how age patterns of U5M are geographically distributed globally, using the direct information collected in Aims 1 & 2 as well as the information collected and adjusted in Aim 3. For this purpose, we will use summary indicators of age patterns such as our k shape parameter. Geographical variation in k will be interpreted in light of background factors known to influence age patterns, such as malaria, HIV and breastfeeding. These qualitative generalizations will be supported more formally by ecological analyses using indicators of malaria prevalence, HIV prevalence, and length of breastfeeding as explanatory variables for a subset of populations where this information is relevant and available. We will also examine joint trajectories in our level vs. shape parameters (U5MR vs. k) over time for settings that have long time series. Following this examination of broad patterns, we will address the following specific questions:

i. How widespread are exceptions to the usual pattern of regular decline in mortality with age from birth to age 5? We will address this question by examining each data set included in Aims 1 & 2, detecting situations in which the risk of death ${}_nM_x$ increases at some point in the 0-5 age range, with statistical tests to assess the significance of such reversals. We will then address this issue more broadly by examining

${}_nM_x$ information in the DHS (adjusted per Aim 3). We anticipate that these unusual patterns, which call for specific interventions targeting the identified ages at reversal, will be found to be more widespread than previously believed. Building on the breastfeeding literature which emphasizes weaning as a critical mechanism in these reversals,[24] we will examine population-level relationships between length of breastfeeding and age at reversal. Population-level breastfeeding patterns will be measured using classic indicators available in HDSS and DHS, including median duration of breastfeeding and prevalence of breastfeeding at various child ages.[30]

ii. Which age groups within the 0-5 age range are most responsive to specific health interventions? We will address this question by focusing on a subset of HDSS and CS that have conducted vaccine trials and other health interventions. We will focus on the following interventions, chosen for their relevance for U5M patterns: Measles vaccination in Niakhar HDSS[59,78] and Bandim HDSS;[79] Oral Rehydration Solution (ORS) in Matlab HDSS;[80] Prevention of Mother-To-Child Transmission (PMTCT) of HIV in Magu HDSS[81] and Africa Centre HDSS;[82] and Insecticide-Treated bed Nets (ITN) for malaria prevention in Kenya (Asembo) CS.[83] Among these interventions, only the Matlab ORS and the Asembo ITN interventions are actual randomized trials, allowing the comparison of ${}_nM_x$ and q(x) in control vs. treatment areas. In spite of their observational designs, the other sites have all provided convincing evidence for the effects of specific interventions on child health through examination of time trends in mortality in relation to the timing of interventions.[14] In either case, we will identify the age ranges, including non-standard age groups, where the estimated intervention effects are the most significant. This comparison will also be made more synthetically by examining joint trajectories in our level vs. shape indicator over time for these sites. Conclusions from this analysis will then be used more broadly for understanding time changes in the age pattern of U5M observed globally in the general analysis described in the introduction of this section.

iii. How do sex differentials in mortality evolve with age within the 0-5 age range in various contexts? This question will be addressed by examining, in the data sets collected in Aims 1 & 2, how the risk of death for males vs. females (${}_nM_x^M/{}_nM_x^F$) varies with age. We will pay particular attention to the amount of sex differential that occurs right at birth (say, during the first 24 hours) in various contexts, because these early ages will be most informative about fundamental genetic and biological differences between males and females.[21,62,63] We will then examine whether sex differentials established right at birth later expand or contract as children get older within the 0-5 age range in various contexts. We will also examine unusual situations in which females experience excess mortality during the 0-5 range. For programmatic purposes, it will be particularly informative to know whether excess female mortality in these contexts occurs as early as the first days or weeks, or whether it emerges at a later age in the life of these girls. Given the small width of our age groups, we expect the amount of sampling error in ${}_nM_x^M/{}_nM_x^F$ to be large. Therefore, most of the analyses in this section will be conducted by pooling the different data sets and extracting broad tendencies using regression analysis. A pooled analysis of the adjusted DHS data will also be used for addressing this question.

C.6. Public-use database. This publicly-available online database will include the information collected in Aims 1 & 2. Only those data sets that were evaluated as high-quality and thus included in the modeling will be integrated. Following the example of the HMD, we will include both the primary aggregate information (births, deaths, and population) as well as the mortality indicators ${}_nM_x$ and q(x), for transparency and reproducibility.

C.7. Timeline and advisory committee. Many of the components of this research are inter-dependent. Therefore this project will be conducted following a specific timeline. The data collection and analysis of HMD countries will be completed by the end of Year 1, with a final version of Model A as the outcome (Aim 1). Information from less-developed settings will be evaluated and gradually integrated into the database during Years 1-3, with final versions of Models A' and B produced by the end of Year 3 (Aim 2). Years 4-5 will be devoted to the indirect estimation of U5M patterns on the basis of DHS and incomplete VR information (Aim 3). Years 4-5 will be also devoted to the analysis of our substantive questions (Aim 4), using the data collected as part of Aims 1 & 2, as well as the DHS and incomplete VR data adjusted with our models as part of Aim 3. The online database will be created during Year 1 with information from HMD countries, and then incremented throughout the project with additional information from the other sources. To facilitate smooth progress, this project will include an advisory committee of 5 members, chaired by Robert Black (Johns Hopkins School of Public Health). More information about the role of this advisory committee is provided in the description of consortium arrangements.

Consortium/Contractual Arrangements

This project brings together an experienced group of researchers spread over several institutions in the US and abroad. These researchers all have expertise that is directly relevant to this application. While they all share broad substantive knowledge in the area of child health and mortality in less-developed countries, each of them brings unique contextual and methodological knowledge, as well as intimate familiarity with specific data sources. This unique expertise justifies the inclusion of **four** different consortium agreements in this application.

(1) Johns Hopkins Bloomberg School of Public Health (Joanne Katz, Co-Investigator)

Joanne Katz will lead a research team at Johns Hopkins that will focus on the analysis of cohort studies (Section C.3.iii of the Research Plan). Dr. Katz has ample experience with these data sets, which she has used previously as leader of a global study on the mortality risks of low birth weight. She was the PI on one of the cohort studies used in this project (in Nepal), and has close collaborative ties with the researchers who led the other cohorts studies. She will lead analysis and dissemination of results of the cohort studies component of this project. She will also contribute to the interpretation and dissemination of results from the pooled analysis of the different project components.

(2) London School of Hygiene and Tropical Medicine, UK (Georges Reniers, Co-investigator)

Georges Reniers will lead a team that will focus on the analysis of HDSS data (Section C.3.ii). Dr. Reniers has ample experience with HDSS data, having been directly involved with several on-going HDSS data collection projects in Sub-Saharan Africa. He will lead the general analysis of the 14 HDSS sites available on iSHARE, as well as the proposed detailed analysis of 3 HDSS sites with which he will directly collaborate (Basse/Faraffeni, Magu and Africa Centre). He will lead the dissemination of the results from these specific project components. He will also work in close collaboration with Drs. Guillot and Pison for joint analysis of all the 8 HDSS sites included in the detailed analyses of Section C.3.ii. He will also contribute to the interpretation and dissemination of results from the pooled analysis of the different project components.

(3) Museum National d'Histoire Naturelle, France (Gilles Pison, Co-investigator)

Gilles Pison will lead a team that will focus on the analysis for 4 HDSS sites situated in West Africa (Niakhar, Mlomp, Bandafassi, Bandim). He has deep knowledge of these data sources, having lead the data collection in one of them and collaborated with the other 3 sites. He will directly interact with HDSS site leaders and disseminate results from these 4 sites. He will also work in close collaboration with Drs. Guillot and Reniers for joint analysis of all the 8 HDSS sites included in the detailed analyses of Section C.3.ii. In addition, he will lead the analysis of the vital registration data from Antananarivo, and disseminate corresponding results. He will also contribute to the interpretation and dissemination of results from the pooled analysis of the different project components.

(4) Johns Hopkins School of Public Health (Li Liu, Co-investigator)

Li Liu will lead a second team at Johns Hopkins, focusing on the analysis of China's MCMSS data. Li Liu has already worked with and published extensively on these data. She already has well-established collaborative ties with the MCMSS staff in China. In addition to the dissemination of the MCMSS results, she will contribute to the interpretation and dissemination of results from the pooled analysis of the different project components.

· United Nations Population Division (Patrick Gerland, Other Significant Contributor)

In addition to these four consortium agreements, this project will involve close collaboration with **Patrick Gerland**, Chief of the Mortality Section of the **United Nations Population Division**. As employee of the United Nations, Patrick Gerland is not able to participate in a formal consortium agreement. However, he will play a key scientific role in this project. He has deep knowledge of all the different data sources used in the project and the methods to analyze them. He will contribute to the design, interpretation and dissemination of results from all the different components of this project. He will also be directly involved with the analysis of the UN database in Sections C.3.iv (Urban Vital Registration) and C.4 (Incomplete Vital Registration).

Synergies between the different teams will be organized as follows. The project will start with a general kick-off meeting in Year 1, gathering the main actors from all these different project components, as well as members of the advisory committee (see below). Due to their common focus on HDSS data, the LSHTM, Museum and Penn teams will work closely together throughout the entire duration of the project, with annual group meetings, as well as a workshop in Year 3 gathering representatives from each participating HDSS site. These teams will also meet frequently via Skype and teleconference. The two Johns Hopkins teams and Dr. Guillot will meet at least twice a year, as facilitated by the short distance between Philadelphia and Baltimore. Drs. Guillot and Gerland will also meet at least twice a year in New York or Philadelphia. Given the many inter-dependencies between the different project components, each team will contribute to the interpretation and dissemination of results from the pooled analyses. A general group meeting in Year 4 will facilitate these interactions. Project participants will also meet frequently at scientific meetings, including the PAA and the IUSSP. Project results will be presented regularly at meetings of the Technical Advisory Group (TAG) of the Inter-agency Group for Child Mortality Estimation (UN IGME), of whom Dr. Guillot is a member and Dr. Gerland a regular attendee in his capacity of Chief of the United Nations Population Division's mortality section, and as UN representative at IGME.

Interactions between the different project components will be further promoted by an **advisory committee** which we have gathered for this project. This advisory committee will be chaired by Robert Black (Johns Hopkins School of Public Health), and includes Jere Behrman (University of Pennsylvania), Simon Cousens (LSHTM), Linda Richter (University of Witwatersrand) and Danzhen You (UNICEF) as members. The charge of the advisory committee will be to periodically evaluate the state of the project as well as advise the project's team on plans for the future. The advisory committee will meet once a year. The advisory committee's first meeting will take place during Year 1 at the Population Studies Center (PSC) of the University of Pennsylvania and will gather all the co-investigators. Subsequent meetings will take place either at the PSC or in some venues offered by meetings such as PAA, IUSSP, and meetings of the TAG, or via teleconference.

Data Sharing Plan

As part of this project, we will create a publicly-available online database containing, in an aggregate format, all of the raw data collected as part of Aims 1 & 2 and used for developing models. We will include both the primary aggregate information (births, deaths, and population) as well as the mortality indicators (${}_nM_x$ and q(x)) derived from it. The database will include a methods protocol, explaining the principles and content of the database, as well as the methodology used for calculating mortality indicators on the basis of the raw information. It will be created during Year 1 with information from HMD countries, and then incremented throughout the project with additional information from the other sources.

This global database will allow the research community to easily access high-quality primary information on under-5 mortality by detailed age for their own research needs, and for further validation and replication of the project's results. Although the data will be provided free of charge to all individuals who request access to the database, we will require users to register and accept a user agreement.

REFERENCES

1 United Nations. *The Millennium Development Goals report 2011*. (United Nations, 2011).

2 UN Inter-agency Group for Child Mortality Estimation. Levels & Trends in Child Mortality: Report 2011 (Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation). (2011).

3 Wang, H. *et al.* Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **384**, 957-979, doi:10.1016/S0140-6736(14)60497-9 (2014).

4 You, D. *et al.* Global, regional, and national levels and trends in under-5 mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the UN Inter-agency Group for Child Mortality Estimation. *The Lancet* **386**, 2275-2286, doi:10.1016/s0140-6736(15)00120-8 (2015).

5 Guillot, M., Gerland, P., Pelletier, F. & Saabneh, A. Child Mortality Estimation: A Global Overview of Infant and Child Mortality Age Patterns in Light of New Empirical Data. *PLoS medicine* **9**, doi:ARTN e100129910.1371/journal.pmed.1001299 (2012).

6 Cantrelle, P. in *Population in African development* (eds Pierre Cantrelle, United Nations. Economic Commission for Africa., & International Union for the Scientific Study of Population.) 33–42 (UIESP & Ordina Editions, 1974).

7 Cantrelle, P. & Leridon, H. Breast feeding, mortality in childhood and fertility in a rural zone of Senegal. *Popul Stud (Camb)* **25**, 505-533, doi:10.1080/00324728.1971.10405821 (1971).

8 Garenne, M. L. *Variations in the Age Pattern of Infant and Child Mortality with Special Reference to a Case Study in Ngayokheme (Rural Senegal)* Ph.D. thesis, University of Pennsylvania, (1982).

9 Behrman, J. R. in *Handbook of Economics and Human Biology* (ed J Komlos) (Oxford University Press, 2016).

10 Leroy, J. L., Ruel, M. & Habicht, J. P. Critical windows for nutritional interventions against stunting. *Am J Clin Nutr* **98**, 854-855, doi:10.3945/ajcn.113.066647 (2013).

11 Maluccio, J. A. *et al.* The Impact of Improving Nutrition during Early Childhood on Education among Guatemalan Adults. *Econ J* **119**, 734-763 (2009).

12 Prentice, A. M. *et al.* Critical windows for nutritional interventions against stunting. *Am J Clin Nutr* **97**, 911-918, doi:10.3945/ajcn.112.052332 (2013).

13 Victora, C. G. *et al.* Maternal and child undernutrition 2 - Maternal and child undernutrition: consequences for adult health and human capital. *Lancet* **371**, 340-357, doi:Doi 10.1016/S0140-6736(07)61692-4 (2008).

14 Ewbank, D. C., Gribble, J. N. & National Research Council (U.S.). Working Group on Effects of Child Survival and General Health Programs on Mortality. *Effects of health programs on child mortality in Sub-Saharan Africa*. (National Academy Press, 1993).

15 Chahnazarian, A., Ewbank, D. C., Makani, B. & Ekouevi, K. Impact of Selective Primary-Care on Childhood Mortality in a Rural Wealth Zone of Zaire. *Int J Epidemiol* **22**, S32-S41 (1993).

16 Preston, S. H. *Mortality patterns in national populations : with special reference to recorded causes of death*. (Academic Press, 1976).

17 Wolleswinkel-van den Bosch, J. H., van Poppel, F. W., Looman, C. W. & Mackenbach, J. P. Determinants of infant and early childhood mortality levels and their decline in The Netherlands in the late nineteenth century. *Int J Epidemiol* **29**, 1031-1040, doi:10.1093/ije/29.6.1031 (2000).

18 Black, R. E. *et al.* Global, regional, and national causes of child mortality in 2008: a systematic analysis. *Lancet* **375**, 1969-1987, doi:10.1016/S0140-6736(10)60549-1 (2010).

19 Rao, C., Adair, T. & Kinfu, Y. Using Historical Vital Statistics to Predict the Distribution of Under-Five Mortality by Cause. *Clinical medicine & research* **9**, 66-74, doi:10.3121/cmr.2010.959 (2011).

20 Liu, L. *et al.* Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet* **379**, 2151-2161, doi:10.1016/S0140-6736(12)60560-1 (2012).

21 Drevenstedt, G. L., Crimmins, E. M., Vasunilashorn, S. & Finch, C. E. The rise and fall of excess male infant mortality. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 5016-5021, doi:10.1073/pnas.0800221105 (2008).

22 Galley, C. & Woods, R. On the Distribution of Deaths During the First Year of Life. *Population: An English Selection* **11**, 35-59 (1999).

23 Huffman, S. L. & Lamphere, B. B. Breastfeeding Performance and Child Survival. *Population and Development Review* **10**, 93-116 (1984).

24 Lamberti, L. M., Fischer Walker, C. L., Noiman, A., Victora, C. & Black, R. E. Breastfeeding and the risk for diarrhea morbidity and mortality. *BMC public health* **11 Suppl 3**, S15, doi:10.1186/1471-2458-11-S3-S15 (2011).

25 Pison, G. & Langaney, A. The Level and Age Pattern of Mortality in Bandafassi (Eastern Senegal): Results from a Small-Scale and Intensive Multi-Round Survey. *Population Studies* **39**, 387-405 (1985).

26 Knodel, J. & Kintner, H. Impact of Breast Feeding Patterns on Biometric Analysis of Infant-Mortality. *Demography* **14**, 391-409, doi:Doi 10.2307/2060586 (1977).

27 Walker, C. L. F. *et al.* Global burden of childhood pneumonia and diarrhoea. *The Lancet* **381**, 1405-1416, doi:10.1016/s0140-6736(13)60222-6 (2013).

28 Cantrelle, P. & Leridon, H. Breast Feeding, Mortality in Childhood and Fertility in a Rural Zone of Senegal. *Population Studies* **25**, 505-533, doi:10.1080/00324728.1971.10405821 (1971).

29 Betran, A. P., de Onis, M., Lauer, J. A. & Villar, J. Ecological study of effect of breast feeding on infant mortality in Latin America. *Brit Med J* **323**, 303-306, doi:DOI 10.1136/bmj.323.7308.303 (2001).

30 Victora, C. G. *et al.* Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. *The Lancet* **387**, 475-490, doi:10.1016/S0140-6736(15)01024-7 (2016).

31 Marinda, E. *et al.* Child mortality according to maternal and infant HIV status in Zimbabwe. *The Pediatric infectious disease journal* **26**, 519-526, doi:10.1097/01.inf.0000264527.69954.4c (2007).

32 Newell, M. L., Brahmbhatt, H. & Ghys, P. D. Child mortality and HIV infection in Africa: a review. *Aids* **18 Suppl 2**, S27-34 (2004).

33 Peden, A. E., Franklin, R. C. & Leggat, P. A. Fatal river drowning: the identification of research gaps through a systematic literature review. *Injury prevention : journal of the International Society for Child and Adolescent Injury Prevention*, doi:10.1136/injuryprev-2015-041750 (2016).

34 Iqbal, A. *et al.* Childhood mortality due to drowning in rural Matlab of Bangladesh: magnitude of the problem and proposed solutions. *Journal of health, population, and nutrition* **25**, 370-376 (2007).

35 Ahmed, M. K., Rahman, M. & van Ginneken, J. Epidemiology of child deaths due to drowning in Matlab, Bangladesh. *Int J Epidemiol* **28**, 306-311 (1999).

36 Black, R. E., Morris, S. S. & Bryce, J. Where and why are 10 million children dying every year? *Lancet* **361**, 2226-2234, doi:Doi 10.1016/S0140-6736(03)13779-8 (2003).

37 Lawn, J. E., Osrin, D., Adler, A. & Cousens, S. Four million neonatal deaths: counting and attribution of cause of death. *Paediatric and perinatal epidemiology* **22**, 410-416, doi:10.1111/j.1365-3016.2008.00960.x (2008).

38 Hill, K. Age patterns of child mortality in the developing world. *Population bulletin of the United Nations*, 112-132 (1995).

39 Anthopolos, R. & Becker, C. M. Global Infant Mortality: Correcting for Undercounting. *World Development* **38**, 467-481, doi:10.1016/j.worlddev.2009.11.013 (2010).

40 Becker, S., Waheeb, Y., ElDeeb, B., Khallaf, N. & Black, R. Estimating the completeness of under-5 death registration in Egypt. *Demography* **33**, 329-339, doi:Doi 10.2307/2061765 (1996).

41 Hall, S. Neonatal mortality in developing countries: what can we learn from DHS data? , (Southampton Statistical Sciences Research Institute, Southampton, 2005).

42 Coale, A. J. & Demeny, P. G. *Regional model life tables and stable populations*. (Princeton University Press, 1966).

43 Coale, A. J., Demeny, P. G. & Vaughan, B. *Regional model life tables and stable populations*. 2nd edn, (Academic Press, 1983).

44 United Nations. *Model life tables for developing countries*. (United Nations, 1982).

45 United Nations. *MortPak-Lite -- the United Nations software package for mortality measurement : interactive software for the IBM-PC and compatibles*. (United Nations, 1988).

46 Galley, C. & Woods, R. Reflections on the distribution of deaths in the first year of life. *Population* **53**, 921-946, doi:Doi 10.2307/1534830 (1998).

47 Lantoine, C. & Pressat, R. New Aspects of Infant-Mortality. *Population* **39**, 253-264, doi:Doi 10.2307/1532294 (1984).

48 Lynch, K. A., Greenhouse, J. B. & Brändström, A. Biometric Modeling n the Study of Infant Mortality: Evidence from Nineteenth-Century Sweden. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* **31**, 53-64, doi:10.1080/01615449809601188 (1998).

49 Manfredini, M. The bourgeois-pichat's biometric method and the influence of climate: New evidences from late 19th-century Italy. *Biodemography and Social Biology* **51**, 24-36, doi:10.1080/19485565.2004.9989081 (2004).

50 Bicego, G. T., Ahmad, O. B. & Macro International. Institute for Resource Development. Demographic and Health Surveys. *Infant and child mortality*. (Macro International, 1996).

51 Blacker, J. G. C., Hill, A. G. & Timaeus, I. in *International Population Conference, Florence, 1985*. 287-298 (International Union for the Scientific Study of Population).

52 Jasseh, M. *Age patterns of mortality within childhood in sub-Saharan Africa* Ph.D. thesis, University of London, (2003).

53 Sullivan, J. M., Rutstein, S. O. & Bicego, G. T. *Infant and child mortality*. (Macro International Inc., 1994).

54 United Nations. *Step-by-step guide to the estimation of child mortality*. (United Nations, 1990).

55 Page, H. in *Population in African development* (eds Pierre Cantrelle, United Nations. Economic Commission for Africa., & International Union for the Scientific Study of Population.) 85–100 (UIESP & Ordina Editions, 1974).

56 Oestergaard, M. Z. *et al.* Neonatal mortality levels for 193 countries in 2009 with trends since 1990: a systematic analysis of progress, projections, and priorities. *PLoS medicine* **8**, e1001080, doi:10.1371/journal.pmed.1001080 (2011).

57 Chu, C. Y., Chien, H. K. & Lee, R. D. Explaining the optimality of U-shaped age-specific mortality. *Theoretical population biology* **73**, 171-180, doi:10.1016/j.tpb.2007.11.005 (2008).

58 Lee, R. D. Rethinking the evolutionary theory of aging: transfers, not births, shape senescence in social species. *Proc Natl Acad Sci U S A* **100**, 9637-9642, doi:10.1073/pnas.1530303100 (2003).

59 Delaunay, V., Etard, J. F., Preziosi, M. P., Marra, A. & Simondon, F. Decline of infant and child mortality rates in rural Senegal over a 37-year period (1963-1999). *Int J Epidemiol* **30**, 1286-1293; discussion 1294-1285 (2001).

60 Lalou, R. & LeGrand, T. K. Child mortality in towns and villages in the Sahel region. *Population* **51**, 329-351, doi:Doi 10.2307/1534584 (1996).

61 Abdullah, S. *et al.* Patterns of age-specific mortality in children in endemic areas of sub-Saharan Africa. *The American journal of tropical medicine and hygiene* **77**, 99-105 (2007).

62 Waldron, I. Sex differences in human mortality: the role of genetic factors. *Soc Sci Med* **17**, 321-333 (1983).

63 Waldron, I. in *Too Young to Die: Genes or Gender?* 64-83 (United Nations, 1998).

64 AbouZahr, C. *et al.* Towards universal civil registration and vital statistics systems: the time is now. *The Lancet* **386**, 1407-1418, doi:10.1016/s0140-6736(15)60170-2 (2015).

65 Aleshina, N. & Redmond, G. How high is infant mortality in central and eastern Europe and the commonwealth of independent states? *Popul Stud (Camb)* **59**, 39-54, doi:10.1080/0032472052000332692 (2005).

66 Guillot, M., Lim, S. J., Torgasheva, L. & Denisenko, M. Infant mortality in Kyrgyzstan before and after the break-up of the Soviet Union. *Pop Stud-J Demog* **67**, 335-352, doi:10.1080/00324728.2013.835859 (2013).

67 HMD. *Human Mortality Database (HMD) [www.mortality.org]*. (University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany)).

68 Wilmoth, J., Zureick, S., Canudas-Romo, V., Inoue, M. & Sawyer, C. A flexible two-dimensional mortality model for use in indirect estimation. *Popul Stud (Camb)* **66**, 1-28, doi:10.1080/00324728.2011.611411 (2012).

69 Barbieri, M. Infant mortality in France. *Population* **53**, 813-838, doi:Doi 10.2307/1534560 (1998).

70 Wang, Y. P. *et al.* Mortality rate for children under 5 years of age in China from 1996 to 2006. *Public Health* **125**, 301-307, doi:10.1016/j.puhe.2011.01.003 (2011).

71 Wang, Y. P. *et al.* Geographical disparities of infant mortality in rural China. *Arch Dis Child-Fetal* **97**, F285-F290, doi:10.1136/archdischild-2011-300412 (2012).

72 Jha, P. *et al.* Prospective study of one million deaths in India: Rationale, design, and validation results. *PLoS medicine* **3**, 191-200, doi:ARTN e1810.1371/journal.pmed.0030018 (2006).

73 INDEPTH Network. *Population and health in developing countries. Volume 1. Population, health, and survival at INDEPTH sites*. 340 (International Development Research Centre, 2002).

74 INDEPTH Network. *INDEPTH model life tables for Sub-Saharan Africa*. (Ashgate, 2004).

75 Katz, J. *et al.* Mortality risk in preterm and small-for-gestational-age infants in low-income and middle-income countries: a pooled country analysis. *Lancet* **382**, 417-425, doi:10.1016/S0140-6736(13)60993-9 (2013).

76 Masquelier, B., Waltisperger, D., Ralijaona, O., Pison, G. & Ravelo, A. The epidemiological transition in Antananarivo, Madagascar: an assessment based on death registers (1900-2012). *Global health action* **7**, 23237, doi:10.3402/gha.v7.23237 (2014).

77 Mikkelsen, L. *et al.* A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. *The Lancet* **386**, 1395-1406, doi:10.1016/s0140-6736(15)60171-4 (2015).

78 Garenne, M. & Aaby, P. Pattern of exposure and measles mortality in Senegal. *The Journal of infectious diseases* **161**, 1088-1094 (1990).

79 Aaby, P., Bukh, J., Lisse, I. M. & Smits, A. J. Measles vaccination and reduction in child mortality: a community study from Guinea-Bissau. *The Journal of infection* **8**, 13-21 (1984).

80 Chen, L. C. *et al.* Mortality Impact of an Mch-Fp Program in Matlab, Bangladesh. *Stud Family Plann* **14**, 199-209, doi:Doi 10.2307/1966412 (1983).

81 Gourlay, A. *et al.* Factors associated with uptake of services to prevent mother-to-child transmission of HIV in a community cohort in rural Tanzania. *Sexually transmitted infections* **91**, 520-527, doi:10.1136/sextrans-2014-051907 (2015).

82 Ndirangu, J., Newell, M. L., Tanser, F., Herbst, A. J. & Bland, R. Decline in early life mortality in a high HIV prevalence rural area of South Africa: evidence of HIV prevention or treatment impact? *AIDS* **24**, 593-602, doi:10.1097/QAD.0b013e328335cff5 (2010).

83 Phillips-Howard, P. A. *et al.* Efficacy of permethrin-treated bed nets in the prevention of mortality in young children in an area of high perennial malaria transmission in western Kenya. *The American journal of tropical medicine and hygiene* **68**, 23-29 (2003).