21

Global Health Estimates: Modelling and Predicting Health Outcomes

Colin Mathers, Dan Hogan, and Gretchen Stevens

1   Introduction

The World Health Organization (WHO) reported that, in 2016, 5.6 million children died before reaching their fifth birthday and almost half of them (46 per cent) died before reaching 27 days. Among the leading causes of their deaths were pneumonia (13 per cent), diarrhoea (8 per cent), congenital anomalies (8 per cent), injuries (6 per cent) and malaria (5 per cent) [1]. Yet about the same time, Countdown to 2030—an independent multi-institutional collaboration that gathers and analyses data on women's and children's health—reported a striking absence of data for causes of child mortality in its 81 high-priority countries. Only 5 of the countries had good quality data for cause of death, 34 had incomplete data and 47 countries had no data at all [2].

So how can WHO make such assertions if the data are so poor? The answer is that WHO uses all available country data on indicators such as these and then makes global and country estimates using statistical modelling. Countdown, on the other hand, 'makes only limited use of predictions and aims as much as possible, to allow country data to speak.' [2]

C. Mathers (✉) • D. Hogan • G. Stevens
Department of Information, Evidence and Research, World Health Organization, Geneva, Switzerland

We review the rationale for estimation in global health and then describe the situations in which models are useful, providing an overview of the major classes of models used. We also discuss issues in assessing the quality and plausibility of statistical estimates and describe recommended guidelines for reporting them. Finally, we examine the relevance of such estimates for countries. We provide references throughout the chapter, to which the reader can refer to learn more about these complex techniques.

2 Rationale and Emergence of Global Health Estimates

Global estimates of health indicators, which are comparable by country, are vital to track progress towards internationally agreed goals and for donors to prioritise their investments. However, the accuracy of these estimates depends on the methods used to create them, and more importantly, the amount and quality of data upon which they are based. Poor data quality or availability means alternative methods can lead to substantially different final estimates, which can cause considerable confusion among global agencies and donors.

A little less than half the deaths in the world are registered with their cause, and national death registration data are only available for four African countries [3]. Useful population level data on incidence or prevalence of disease and injury are even less available. Instead international agencies and academics use statistical models to prepare estimates of key health indicators that are comparable across countries and/or time. The agencies derive these *global health estimates* using reported or published data from multiple national sources, such as civil registration, health facilities and population surveys. Estimates are valuable in generating overviews of the global health situation and emerging trends, and for reporting on country and global progress towards the Millennium Development Goals (MDGs) and now the Sustainable Development Goals [4].

Starting from the 1950s, and with increasing scope and regularity since the 1990s, the United Nations (UN) and its specialised agencies, such as the WHO and the UN Children's Fund, have published annual global and country health estimates for major demographic and health indicators based on data reported by member states. Within the last decade, the Institute for Health Metrics and Evaluation (IHME), funded by the Bill & Melinda Gates Foundation, has also published annual updates of comprehensive global burden of disease (GBD) statistical time series based on available data for 195

countries and territories, with sub-national estimates for a growing number of countries [5].

The statistical models used by various groups vary widely and dramatic expansion of computing and storage capacity has facilitated increased technical complexity. For example, the WHO estimated that in 2010, there were 655,000 malaria deaths worldwide, with under 100,000 in those aged five years and over [6]. The IHME estimated equivalent figures of 1.24 million malaria deaths, with more than half a million occurring in those aged five years and older [7]. Differences in interpretation of data, inclusion criteria and methodologies have led to publication of very different values for the same indicator. This can have serious consequences for individual countries. Depending which estimate they think to be more reliable, global donors may assign funding and evaluate progress differently. This situation has heightened calls from international agencies, policymakers and researchers for more transparency and replicability of methods. Some national policymakers and data producers question the need for such techniques, preferring to use their national statistics, where they are available.

3   Why Model?

Raw health data derived from primary data collection are often reported as direct tabulations of counts or transformed into indicators, such as rates or ratios without any adjustments or corrections. These statistics may not be accurate, representative of the population of interest, or comparable. Drawing comparisons between populations can also be complicated by differences in data definitions and measurement methods. Some countries may have multiple sources of data for the same population-time period, but more often data are not available for every population and year. Box 21.1 describes some common sources of bias. To overcome these issues, statisticians use analytic methods, such as mathematical and statistical models, to produce unbiased estimates that are representative and comparable across populations and/or time.

The types of modelling used vary in sophistication, but share the goal of addressing some or all of these challenges. We describe below the key situations in which modelling is useful.

To Improve Accuracy and Comparability of Data. One major purpose of statistical modelling is to process raw data to improve its accuracy and comparability. The application of weighting factors to data collected in a cluster sample survey is a form of modelling to improve representativeness.

> **Box 21.1 Common Sources of Bias in Model Input Data. Adapted from the GATHER Statement [8]**
>
> *Inconsistent case definitions or diagnostic criteria:* Health data often identify persons who test positive for a particular case definition. Case definitions may vary by data source, limiting their comparability. Assessors' qualifications may vary, which can lead to differences in ascertained prevalence. In addition, laboratory protocols may change over time, reducing comparability even when case definitions have not changed. Changes in sensitivity or specificity of detection methods can have an important effect on case identification, as can decisions about whether to adjust for sensitivity or specificity.
>
> *Self-report biases:* With some survey instruments, systematic biases can arise from difficulties in obtaining accurate responses from survey respondents. Examples of self-report biases include recall bias or social desirability bias. Self-reports of prior diagnosis often underestimate the true incidence or prevalence since some cases do not interact with the health system or are not diagnosed. These biases may vary systematically by populations and over time.
>
> *Incomplete population-based surveillance:* Surveillance and registration systems designed to capture all events in a population are often incomplete. It may be difficult to quantify levels of completeness for events such as infectious disease incidence. For other types of events, demographic techniques or capture-recapture techniques may allow estimation of completeness.
>
> *Non-representative population bias:* Some data types are collected for a subset of the general population by design, for example when data are collected from clinic attendees or samples of volunteers, or when data pertain to urban or rural groups only. Health status and health determinants may differ systematically between these selected populations and the general population.

Incompleteness of surveillance or registration data is an important source of bias that must be addressed, and poses a challenge because completeness cannot be assessed using primary data alone and data from other sources is also needed.

Analysts may address bias resulting from definitional and measurement issues a priori by adjusting the data before statistical modelling, drawing on external information. For example, it is possible to adjust the prevalence of hearing loss measured using different loudness decibel thresholds to a common threshold using a known or assumed relationship between threshold and cumulative prevalence. Alternately, adjustment of data using different measurement strategies may be carried out statistically in the model. This is known as *cross-walking* to a standard definition. For example, for multiple hearing loss surveys with different thresholds, analysts can use statistical models to estimate the relationships between thresholds and prevalence and produce estimates for a standard set of thresholds for all the populations.

A striking example of the challenge of comparability comes from Malawi. The 2001 National Micronutrient Survey found that 59 per cent of pre-school

age children had vitamin A deficiency, based on a measure of serum retinol [8]. Surveys in 2009 [9] and in 2015–2016 [10] found prevalence of 40.1 per cent and 3.6 per cent using a different measure—retinol binding protein. Development partners and funding agencies need to know whether the trend indicates programmatic success, or if it is simply due to the change in diagnostic methods, in order to decide how to allocate future funds. This example, and others, highlights the importance of understanding and communicating why and how global health estimates are produced, and their levels of uncertainty.

To Synthesise Data from Multiple and Overlapping Sources. Statistical modelling can also be used to generate comparable and consistent indicator values across populations and/or time—based on all the data which meet inclusion criteria. For example, some countries have multiple sources of data on under-five mortality, such as from the census and household surveys [11]. Synthesising data makes use of all existing information of sufficient quality, thereby avoiding the arbitrariness of an analyst picking the *best* single data source, which is challenging given the presence of measurement error. This approach is similar to estimating a treatment effect through a meta-analysis of several randomised trials as opposed to picking the treatment effect from only one of the trials.

To Fill Data Gaps in time series and project to a common target year or range of years. For most types of raw data, the date of most recently available data varies across populations. Because analysts usually want to estimate trends to a common recent year across all populations, they include a projection component in the model. These imputation methods often borrow information from neighbouring data, which could be, for example, from countries in the same region or other time points in a country's primary data series. Analysts may also seek to improve imputations and projections by including predictor variables in the model that correlate with the quantity of interest (these are known as *covariates*).

To Estimate Quantities that Cannot Be Directly Measured. When it is difficult or costly to measure a health outcome, it may be more feasible to measure intermediate outcomes, and then use a model to extrapolate to the outcome. Such models usually involve mathematical modelling of the causal chain. For example, WHO has based its estimates of measles mortality on estimates of measles cases multiplied by separately estimated case fatality rates [12].

To Evaluate Large-scale Public Health Interventions when a randomised controlled trial is not possible for ethical or practical reasons [13]. Investigators observe trends in the outcome of interest with the programme in place and

develop a counterfactual model to estimate the outcomes in the absence of the programme. This approach can also be used to assess the potential impact and cost-effectiveness of proposed interventions.

To Forecast Indicators for a standard time frame (base year to latest target year) using a forwards (and sometimes backwards) projection component. In some cases, the main aim of the modelling is longer range projection or forecasting. These types of models fall into two main classes: (1) deterministic covariate-driven projections that are usually *scenario-based* [14, 15]. These allow for modelling of alternate future policies or interventions through covariates or other modifiable parameter assumptions; and (2) statistical forecasts using time series projection techniques to extrapolate historical trends [16]. Hybrid models combine stochastic time series projections with covariate drivers and multi-level modelling [17].

We distinguish in the following sections between statistical models, which describe associations between variables, and mathematical models which postulate a causal pathway [16]. We describe statistical modelling in more detail as analysts use these most frequently to make global estimates.

4   Mathematical Modelling

Mathematical models set up a theoretical framework that represents and quantifies the causal pathways and mechanisms linking determinants and health outcomes. These types of models make predictions of health outcomes (which may be difficult to measure) based on parameter estimates derived from various data sources. An example of a simple mathematical model used in the first GBD study [18] was the DISMOD I model. This specified the basic relationships between incidence, prevalence, remission, case fatality and mortality in terms of a set of four interlinked differential equations (see Fig. 21.1).

Natural history models are commonly used to estimate mortality from various infectious diseases. Recent examples include the UNAIDS HIV Spectrum model [19], the WHO measles mortality model [12] and a rabies mortality model [20]. Garnett et al. [13] give a range of examples of more sophisticated mathematical models, which lend themselves to programme evaluation by modelling the consequences to the final outcome variables of variations in intermediate parameters such as intervention coverage or case fatality. The Comparative Risk Assessment methodology developed by WHO in the early 2000s [21] also uses a mathematical modelling framework. The model assesses the change in population mortality outcomes associated with counterfactual
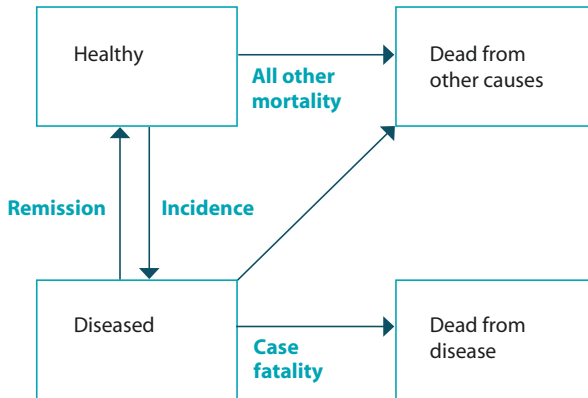
Fig. 21.1 The DISMOD 1 conceptual disease model [18]. The four boxes for prevalence and deaths are linked by four transmission hazards

exposure distributions for risk factors in order to assess the mortality attributable to current and past risk factor exposures.

5   Statistical Modelling

Statistical models estimate or predict outcome indicators using empirical data on the outcome as well as on correlated variables, or covariates. Statistical models commonly use regression techniques, identifying a functional form which fits the data, and which gives an adequate summary of the variation in the data [22]. Whereas explanatory modelling seeks to accurately characterise relationships between variables in the data, prediction modelling aims only to predict outcomes.

5.1   Methods of Estimation

Use of Covariates

Statistical models may estimate and use the correlation between data observations and covariates to improve predictive validity. This approach is frequently used to generate values for indicators in settings with no or very limited primary data on the outcome of interest, for example, levels and trends in maternal mortality [23] and other causes of death. Issues of causality are irrelevant for these types of models and users must be warned not to interpret the associations in causal terms. Analysts should not restrict the choice of covariates to those believed to be causal, as the aim is accurate prediction.

When using covariates, there is a danger that estimated trends reflect changes in covariates rather than changes in the estimated outcome indicator, particularly when there is little outcome data. For example, models to predict maternal mortality often include covariates such as gross domestic product (GDP) per capita which can vary depending on commodity prices. Rising GDP per se may have next to no impact on maternal health over short time periods but a model that includes GDP as a covariate will predict reductions in maternal mortality.

Inclusion of *data type* covariates in a regression is a common strategy when there are datasets or countries where data are available according to several definitions. It is possible to cross-walking to the preferred definition by including indicator variables for each alternate data type in a regression analysis and then setting data type to the preferred type for producing regression estimates. Alternately, it may be more convenient to do the cross-walking as a pre-processing step based on a separate regression analysis. An example is a recent study of diabetes mellitus prevalence which included some data sources that identify diabetes using HbA1c measurements and others that measure fasting plasma glucose [24].

An example of the use of covariates for both prediction of levels and trends and for cross-walking between two data types is the model used by WHO for estimating national homicide rates across countries [25]. After cross-validation, the final model included covariates for alcohol-drinking pattern, gender inequality index, per cent of the population living in urban areas, proportion of the population that are 15–30-year-old males, religious fractionalisation and infant mortality rate. An additional covariate for data type distinguished data from criminal justice and police systems from that derived from death registration, and adjusted for the differences between them.

Frequentist Versus Bayesian Estimation Methods

Frequentist statistical methods are based on interpretation of probabilities as objective summaries of repeated trials of the same process. *Frequentist* statistical modelling methods (such as ordinary least squares regression) rely on maximising a likelihood function which summarises the conditional probability of the actual observations as a function of the parameters to be estimated. In contrast, the Bayesian paradigm treats probabilities as subjective assessments based on prior knowledge (prior probability distributions) which are updated in the light of observed data [26].

Bayesian methods generally allow the fitting of more complex and flexible models, that seek to make many internal adjustments, enable more appropriate uncertainty characterisation and avoid the approximations required for many classical frequentist methods. These methods require greater computation than frequentist methods. With increasing computing power, Bayesian methods have become tractable for virtually all parametric methods and are being increasingly adopted for global health modelling, for example, UN agencies now use Bayesian methods to monitor child and maternal mortality [27, 28].

5.2   Types of Model

In Sect. 3, we identified a number of objectives for using statistical modelling; here we examine some of the main features and uses of the relevant modelling approaches.

Modelling to Produce Smooth Estimates Across Multiple Observations

Complex curve smoothing or time series projections allow flexibility in curve fitting using multiple and sometimes overlapping data inputs. For example, the UN Inter-agency Group on Mortality Estimation (UN-IGME) formerly used a *loess* regression method to estimate trends in child mortality for a country across a standard time period [28]. This method only used country-specific data to interpolate and extrapolate a smooth curve for a single population. UN-IGME now models time trends using Bayesian bias-adjusted B-splines which allow more objective curve fitting than loess regression [27]. We describe this example to illustrate the increasing sophistication of current statistical modelling.

Using the B-splines model, the UN-IGME estimates a best fit trend line for the under-five mortality rate (U5MR) based on multiple observations from multiple surveys at uneven intervals (see Fig. 21.2) [30]. The B-splines model includes a data model which simultaneously adjusts for statistically estimated biases for each type of measurement technique (such as indirect birth history vs. direct birth history). For example, if on average (across all country data) indirect birth history observations were 10 per cent lower than the final U5MR estimates based on all types of data, then the data model will apply an upward adjustment to the indirect birth history observations when estimating the final curve for U5MR. This means that the final
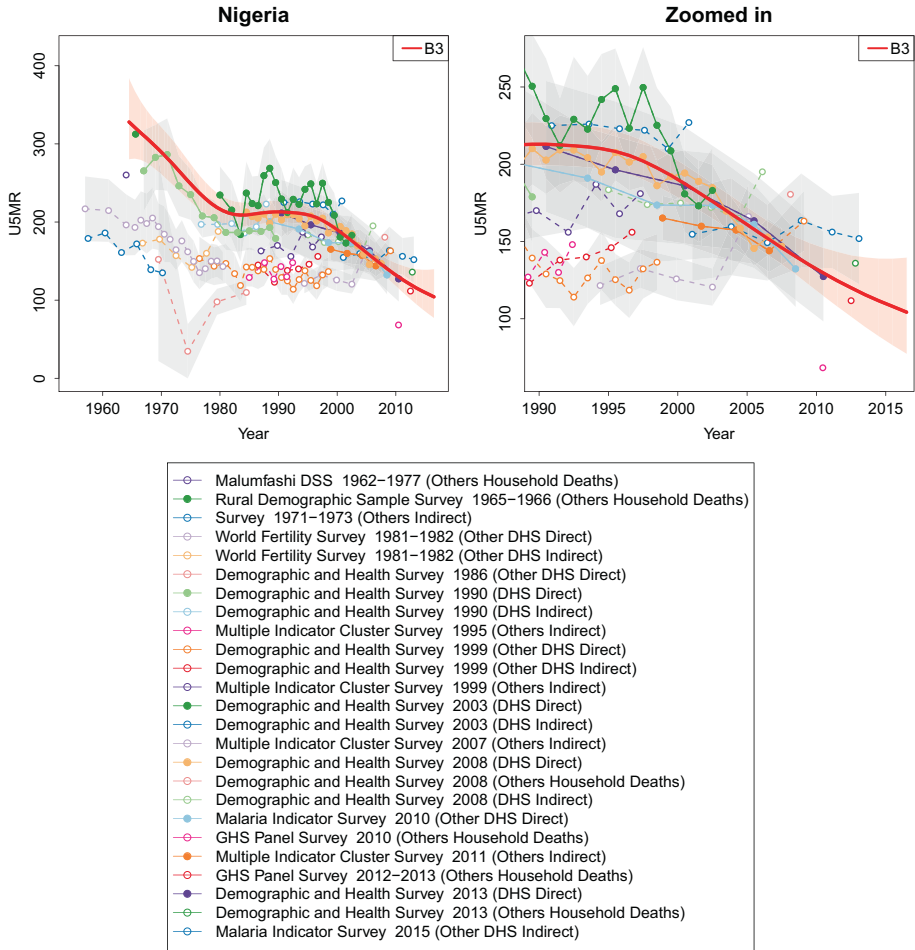
Fig. 21.2 Under-five mortality rates for Nigeria, 1955–2016. (Source: United Nations [30]). Empirical data from surveys and censuses included in the statistical analysis are shown as solid lines with symbols, data excluded on grounds of low quality shown as dashed lines, UN-IGME estimated time series shown as bold red line with 90 per cent uncertainty range

estimated U5MR curve for a country depends on the data for all countries, not just that specific to the country in question. It can also mean that if a country only has observations from biased data sources, the final estimates may lie entirely outside the original, raw data observation (often difficult to explain to users of the statistics).

Multi-level Modelling to Improve the Quality and Stability of Estimates Based on Relatively Sparse Data

Multi-level or hierarchical regression models allow for simultaneous modelling of parameters that vary at more than one level (e.g. country, region and world). Modelling parameters hierarchically allows data from other countries within a region, and in other regions, to inform estimation for countries with poor or missing data. In non-hierarchical regression models, a group *dummy variable* could be included to estimate the variation between groups as a *fixed effect*. Hierarchical models also permit the inclusion of random effects, which allow the model to share information from higher levels of the hierarchy to a greater extent when data from lower levels are poor [22]. We describe an application of hierarchical modelling to children's height and weight to illustrate how the method is typically used for global health estimates. Paciorek et al. used a Bayesian hierarchical model to estimate distributions of height-for-age and weight-for-age by place of residence (urban or rural) for 141 countries over a 35-year period [31]. The estimated values for each country-year were informed by data from the country-year itself, if available, and by data from other countries, especially those in the same region. The authors of the study explained that 'The hierarchical model shares information to a greater extent when data are non-existent or weakly informative (e.g., because they have a small sample size), and to a lesser extent in data-rich countries and regions.'

Complex Predictive Models to Interpolate and Extrapolate Outside the Available Data

Most statistical models use more than one of the techniques outlined above, including use of time-varying covariates, a multi-level structure and a temporal smoothing technique. For example, to estimate maternal mortality trends by country, the UN system uses a multi-level Bayesian regression model with time series modelling and covariates and random effects [28]. The Maternal Child Health Epidemiology Estimation collaboration with WHO uses a multi-nomial regression model, with covariates and fixed effects, that simultaneously models a complete set of cause-of-death fractions [32]. Other examples of complex statistical models include use by the IHME of Gaussian process regression to borrow strength and smooth across space and time [33]. Use of these statistical imputation and prediction methods along with predictive covariates now enables relatively sparse data to become *big estimates* with health indicators imputed to detailed spatial-temporal grids, for example, 5 km x 5 km grids for the world over 15 years [34].

5.3   Appropriateness of Model Frameworks and Validation Methods

Validation of predictive models differs from validation of explanatory models. Analysts validate explanatory models by examining whether their structure adequately represents the data and if the model fits the data. For example, validation of an explanatory model would examine whether addition of extra covariates to the model, transformations of covariates or additional nonlinear terms significantly increase its explanatory power. Model fit is assessed using goodness-of-fit tests and model diagnostics such as residual analysis [35].

In contrast, for predictive modelling where observations are missing for some populations or time periods, the focus of validation is on the ability of the model to predict missing data. This usually involves withholding some of the data from the model fitting and then testing the accuracy of the model predictions against the withheld data, known as *out-of-sample predictive validation* or *cross-validation* [36, 37]. Predictive validity depends on the question being asked and the nature of the data to which it is being applied, so there is no standard metric for evaluation of model performance. For example, a model focused on estimation of the outcome for all countries for a target year that falls outside the dataset will require the model to be particularly good at out-of-time predictions. If the focus is on prediction for countries with no primary data, this requires that the model predicts well out-of-sample across countries. For assessing the predictive validity of cause-of-death models used in the GBD 2010 study, the withheld data consisted of a mix of five types of missing data: countries with no data; countries with missing data years within the available data; countries with missing data years at earlier time periods; countries with missing data for later time periods; and countries with data missing for some age groups [38].

Advances in other disciplines [38, 39, 40] have found that an ensemble modelling approach may give better predictive validity than any single model. Recent modelling in the global health field has also made use of ensemble models that are the weighted combinations of different models [37–39]. Such ensemble modelling typically requires two sets of withheld data for validation. The first set is used to assess the predictive validity of the individual models, and the second set is used to assess and maximise the predictive validity of the ensemble average.

6   Understanding, Assessing and Using Statistical Estimates

Increasing complexity of models being used for health estimates and increasing concerns about the transparency and replicability of modelled results led WHO to assemble a working group in 2014 to define and promote best practice in reporting health estimates. This resulted in a consensus statement and reporting list known as Guidelines for Accurate and Transparent Health Estimates Reporting (GATHER), published in 2016 simultaneously in the Lancet and PLoS Medicine [41, 42].

GATHER defines best reporting practices for global health estimates. GATHER comprises a checklist of 18 items that are essential for the best reporting practice. Key items in this checklist include information on all included data sources and their main characteristics, a detailed description of all steps of the analysis, types of uncertainty quantified and the methods used, how to obtain the analytic or statistical source code used, reasons for changes if updating an earlier set of estimates and a discussion of the modelling assumptions and data limitations that affect interpretation of the estimates. More details are available on the GATHER website [41].

GATHER provides an achievable standard for reporting health estimates, but there are many challenges in implementation. Full documentation of a study typically requires lengthy technical appendices, and ensuring open access to input data and computer code implies an additional reporting burden when publishing estimates. A clear description of the methods and fair discussion of limitations are important for understanding estimates, but are not easy to provide or verify.

6.1   Uncertainty Estimation

Quantifying uncertainty around modelled health estimates—typically by calculating and reporting uncertainty intervals—was considered by the GATHER working group to be a necessary component of reporting results, encouraged by GATHER [42]. Uncertainty ranges provide users with an understanding of the precision of the estimates, and are critical for making comparisons. However, the inclusion of the main sources of uncertainty usually requires substantial statistical expertise and computing power.

Potential sources of uncertainty include stochastic errors, sampling error, and non-sampling errors (resulting from measurement errors, missing data, errors in coverage and other systematic biases). They also include error in

model covariates, parameter uncertainty, model specification uncertainty, fundamental uncertainty and uncertainty arising from various data transformation steps and externally derived parameters [41]. In practice, most quantitative uncertainty estimates reflect only a subset of all possible sources of uncertainty in the estimates.

There is no established methodology for estimating some types of uncertainty. Analysts may use different methods, including developing new methods, or ignore the source of uncertainty and acknowledge this as a limitation of their analysis. In many cases, the data and information needed to quantify uncertainty do not exist (e.g. some sources of uncertainty may be unknown, or impossible to measure). This means that some modelling approaches have wider uncertainty than others, simply because the former may be capturing more sources of uncertainty. In general, accounting for multiple sources of uncertainty—and correctly reflecting these sources in resulting estimates—is more straightforward when Bayesian approaches are used. Uncertainty in values of covariates such as average income per capita or in denominator values such as population estimates is also typically not available and not included in quantitative uncertainty ranges for modelled health indicators.

6.2   Uncertainty Versus Sensitivity Analysis

All estimation processes involve assumptions, including about inclusion criteria for data and the functional form of a model. Some analysts may choose to use sensitivity analysis to assess the degree to which the final values of the estimates depend on these assumptions. If the sensitivity analysis suggests that various analytical approaches produce similar estimates, this lends credibility to the estimates and strengthens the results. If, on the other hand, the sensitivity analysis suggests that the estimates are highly dependent on the modelling approach or the data inclusion/exclusion criteria, this encourages readers to examine carefully the analytical assumptions, and may help to inform future research.

6.3   Transparency, Replicability and Complexity

Transparency is at the heart of controversies about global health estimates. The more diverse the available raw data, the modifications to the raw data and the statistical models, the more difficult it is for an external party to understand and replicate the findings. Also, analysts need to carefully consider the benefits of more model complexity. If their resulting estimates are

similar, a simpler model that others can more easily replicate and use is likely to be more effective than a more complicated model that can only be run and understood by a few individuals. Furthermore, while it is of value to offer greater technical documentation as per the GATHER guidelines, this alone may not be enough to inform users about appropriate interpretation. Many users may lack the technical understand the methods and their limitations. GATHER also requires a plain-language description of methods and a fair discussion of limitations, however, researchers and users may disagree about what constitutes plain language, and a frank discussion of limitations may be perceived as damaging the credibility of the estimates.

6.4   Communicating Estimates

In many cases, estimates that are largely imputed are not clearly flagged as such to users and full documentation of statistical methods is difficult to obtain or understand. Ideally, estimates are presented with uncertainty ranges or confidence intervals, but the utility of these uncertainties is often not clear to users and decision-makers. For its estimates of mortality by cause, WHO uses a four-colour coding system to indicate the strength of the underlying data and whether models and data have used mainly country-specific data or borrowed strength from other countries or covariates. More discussion is needed as to whether and how uncertainty ranges can contribute to better communication and use of estimates.

7   Divergent Health Statistics: Exposing the Limitations in Modelling and Data

For many health indicators multiple global health estimates are now available: one from the UN system and others from academic institutions. This can be of concern to international users such as donor agencies and to national governments [43, 44, 45].

Both WHO and IHME publish regular updates of estimated global deaths by cause [46, 47]. The WHO cause-of-death estimates draw on WHO and UN agency/inter-agency statistics and put them into a consistent comprehensive context for all causes. They also draw on death registration data and IHME GBD analyses for causes/countries without death registration data and where the UN system has not invested in detailed estimates. Over time, there has been some convergence between GBD and WHO estimates for some causes,

though major differences remain in some areas such as adult malaria mortality and tuberculosis cases [39].

The WHO estimates use the latest UN Population Division life tables for total deaths by age and sex, with some adjustments for high HIV countries and for countries with relatively complete death registration. The GBD model life tables differ significantly in some aspects. For example, the GBD 2015 estimated 8.0 million deaths for 2015 for the WHO African region compared with the UN providing an estimate of 9.1 million deaths. The most recent GBD update used IHME birth estimates substantially lower than the UN estimates, resulting in more than 10 per cent reduction in estimated child deaths compared with the latest UN inter-agency estimates [11]. Future revisions of the IHME GBD study will use IHME estimates of population numbers, likely resulting in additional divergences in numbers of deaths.

Like international rankings, dissonant health statistics can cut both ways. In some cases, they can be demoralising, undermining the ability or will to invest in programmes whose success is not yet reflected in global statistics. In other cases, they have led to national debate and greater national investment in data collection and analysis [48]. A critical lesson that has emerged from such debate is the need for much greater dialog between agencies carrying out global estimates and national authorities. They need to discuss the data limitations and biases being addressed through the global modelling process, and to develop a shared understanding of the strengths and limitations of both the input data and the estimates derived from global statistical models.

8   Are Modelled Estimates Helpful for Health Decision-Makers and Consumers?

Users of health statistics have different data needs. The perceived credibility and utility of different kinds of statistics vary significantly by user. National and sub-national data users often prefer empirically measured data that can inform decision-making at national and sub-national levels. Such users are less concerned about comparability with other national estimates or international standards. By contrast, global users, including international agencies, donors and development partners value estimates that are comparable across countries and over time. This translates into variations in the types of statistics that are considered most credible at different levels of governance. This, in turn, affects the likelihood that statistics will be used to inform policy.

Ways in which global estimates can be useful for countries include: comparative analyses of country values (benchmarking with peer countries); progress monitoring and reporting for global and regional goals and targets; reporting to donors and development partners, for example, for performance-based grants; and for estimating completeness and accuracy of empirical reported data.

One challenge for some users of global health estimates is that each revision typically involves a complete re-estimation of the whole time series rather than adding new values for recent years. In some cases, such as child mortality rates that incorporate survey responses with 15 or more years of historical recall, these changes to the time series are based on new empirical data and are explainable. However, in other cases, the data may remain the same, but changes to the estimation methods lead to substantial differences in the estimated series. This can cause confusion, for example, if baseline estimates change, with implications for the speed—and even the direction—of time trends and shifts away from or potentially even over final targets. While the differences usually fall within margins of uncertainty, it can be difficult to explain these changes to policymakers.

Another relevant factor is that health statistics are often used for political purposes. Globally produced statistics that differ substantially from country-reported data can be seized upon for political purposes. Governments may use favourable estimates to rally support for current policies. Conversely, unfavourable estimates bolster political opposition and civil society criticisms of the government. This makes it all the more important to ensure greater shared understanding of the reasons for global modelling adjustments to raw input data.

Global health estimates do not replace the need for countries to collect reliable, accurate and regular empirical data. However, using estimates to fill in missing data can mislead users into thinking the empirical data are available, and reduce pressure to improve information systems. Production of estimates remotely, using complex modelling techniques, may also undermine country understanding and ownership of their indicators. And in an era of global target setting, there is a danger that predicted statistics may be used for the evaluation of progress. The production of estimates should go hand-in-hand with development of tools and methods that build capacity in countries for data generation, analysis and interpretation.

9   Conclusion

In principle, it is possible to track events such as birth and death, cancer incidence and some types of injury by complete registration or surveillance systems. But most population health indicators will continue to be based on data from health information systems (which do not capture all events within populations), epidemiological studies and regular or irregular sample surveys that may rely on self-report by respondents. Synthesis of population indicators from such data will continue to require statistical modelling, though model complexity could diminish if countries adopt universal standards for regular representative sample surveys.

The increasing demand for health data for monitoring the Sustainable Development Goals [49], across a much broader range of health issues than the MDGs, may result in additional investment in good quality population health data, but this will take considerable time to achieve. The world will continue to rely on statistical modelling for almost all health indicators at global and regional levels for many years to come.

Key Messages

- Demand is high for global health statistics that are comparable across countries and time, to be used for priority setting and to monitor health systems performance.
- Reported statistics can be limited by non-standard data definitions, incompleteness and other sources of bias.
- International agencies and academic institutions use statistical and mathematical models to estimate comparable global health statistics.
- Model complexity is increasing as statistical methods advance and computing power increases.
- Good practice reporting principles are available to increase transparency and replicability of methods.

References

1. World Health Organization Global Health Observatory. Causes of child mortality. 2016 [Available from: http://www.who.int/gho/child_health/mortality/causes/en/
2. Countdown to 2030 Collaboration. Countdown to 2030: tracking progress towards universal coverage for reproductive, maternal, newborn, and child

health. Lancet 2018;391:1538–48. Available from: http://dx.doi.org/10.1016/S0140-6736(18)30104-1

3. World Health Organization. World health statistics 2017. Geneva, Switzerland: World Health Organization. 2017 [cited 2018 25th January]. Available from: https://www.who.int/gho/publications/world_health_statistics/2017/en/

4. World Health Organization. Health in 2015: From Millennium Development Goals to Sustainable Development Goals. Geneva, Switzerland: World Health Organization. 2015 [cited 2018 25th January]. Available from: https://www.who.int/gho/publications/mdgs-sdgs/en/

5. Murray CJL, Lopez AD. Measuring global health: motivation and evolution of the Global Burden of Disease Study. Lancet. 2017 Sep;390(10100):1460–4. Available from: http://dx.doi.org/10.1016/s0140-6736(17)32367-x

6. World Health Organization. World Malaria Report 2011. Geneva, Switzerland: World Health Organization; 2011 [cited 2018 5th November]. Available from: http://www.who.int/malaria/world_malaria_report_2011/en/

7. Murray C, Rosenfeld L, Lim S, Andrews K, Foreman K, Haring D, et al. Global malaria mortality between 1980 and 2010: a systematic analysis. Lancet. 2012;379:413–31. Available from: http://dx.doi.org/10.1016/S0140-6736(12)60034-8.

8. Centers for Disease Control and Prevention (CDC), Ministry of Health (Malawi), National Statistical Office of Malawi, United Nations Children's Fund (UNICEF). Malawi National Micronutrient Survey 2001. Lilongwe, Malawi; 2003.

9. Department of Nutrition, HIV and AIDS in the Office of President and Cabinet (DHNA-OPC), Ministry of Health (MOH), National Statistics Office (NSO), United Nations Children's Fund, Centers for Disease Control and Prevention (CDC). The National Micronutrient Survey 2009. Lilongwe, Malawi; 2011.

10. National Statistical Office (NSO), Community Health Sciences Unit (CHSU) [Malawi], Centers for Disease Control and Prevention (CDC), Emory University. Malawi Micronutrient Survey 2015–16: Key indicators report. Lilongwe, Malawi; 2016 [cited 2018 5th November]. Available from: https://dhsprogram.com/pubs/pdf/FR319/FR319m.pdf

11. United Nations Inter-agency Group for Child Mortality Estimation. Levels and trends in child mortality. 2017 [cited 2018 25th January]. Available from: www.childmortality.org

12. Simons E, Ferrari M, Fricks J, Wannemuehler K, Anand A, Burton A, et al. Assessment of the 2010 global measles mortality reduction goal: results from a model of surveillance data. Lancet. 2012;379(9832):2173–8. Available from: http://dx.doi.org/10.1016/S0140-6736(12)60522-4

13. Garnett GP, Cousens S, Hallett TB, Steketee R, Walker N. Mathematical models in the evaluation of health programmes. Lancet. 2011;378(9790):515–25.

14. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Medicine. 2006 Nov 28;3(11):e442. Available from: http://dx.doi.org/10.1371/journal.pmed.0030442

15. Frederick S. Pardee Center for International Futures. The International Futures (IFs) modeling system, version 7.31. University of Denver. [cited 2018 24th January]. Available from: http://www.pardee.du.edu/

16. Kontis V, Bennett JE, Mathers CD, Li G, Foreman K, Ezzati M. Future life expectancy in 35 industrialised countries: projections with a Bayesian model ensemble. The Lancet. 2017 Apr;389(10076):1323–35. Available from: http://dx.doi.org/10.1016/s0140-6736(16)32381-9

17. Girosi F, King G. Demographic forecasting. Princeton, USA: Princeton University Press; 2008.

18. Murray C, Lopez AD. The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020: World Health Organization. 1996 [cited 2018 25th January]. Available from: http://apps.who.int/iris/handle/10665/41864

19. Stover J, Andreev K, Slaymaker E, Gopalappa C, Sabin K, Velasquez C, et al. Updates to the SPECTRUM model to estimate key HIV indicators for adults and children. AIDS. 2014 Nov;28:S427–S434. Available from: http://dx.doi.org/10.1097/qad.0000000000000483

20. Hampson K, Coudeville L, Lembo T, Sambo M, Kieffer A, Attlan M, et al. Estimating the global burden of endemic canine rabies PLoS Neglected Tropical Diseases. 2015 Apr 16;9(4):e0003709. Available from: http://dx.doi.org/10.1371/journal.pntd.0003709

21. Ezzati M, Lopez AD, Rodgers A, Murray CJL. Comparative quantification of health risks. Global and regional burden of disease attribution to selected major risk factors: World Health Organization. 2004 [cited 2018 25th January]. Available from: http://apps.who.int/iris/bitstream/10665/42792/1/9241580348_eng_Volume1.pdf

22. Gelman A, Hill J. Data analysis using regression and multilevel hierarchical models. New York, NY, USA: Cambridge University Press; 2007.

23. Wilmoth JR, Mizoguchi N, Oestergaard MZ, Say L, Mathers CD, Zureick-Brown S, et al. A new method for deriving global estimates of maternal mortality. Statistics, Politics, and Policy. 2012 Jan 13;3(2). Available from: http://dx.doi.org/10.1515/2151-7509.1038

24. NCD Risk Factor Collaboration. Effects of diabetes definition on global surveillance of diabetes prevalence and diagnosis: a pooled analysis of 96 population-based studies with 331,288 participants. Lancet Diabetes & Endocrinology. 2015 Aug;3(8):624–37. Available from: http://dx.doi.org/10.1016/s2213-8587(15)00129-1

25. Butchart A, Mikton C. Global status report on violence prevention. 2014 [cited 2018 25th January]. Available from: http://www.who.int/violence_injury_prevention/violence/status_report/2014/en/

26. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. Third ed. Boca Raton, USA: CRC press; 2014.

27. Alkema L, New JR. Global estimation of child mortality using a Bayesian B-spline Bias-reduction model. The Annals of Applied Statistics. 2014 Dec;8(4):2122–49. Available from: http://dx.doi.org/10.1214/14-aoas768

28. Alkema L, Zhang S, Chou D, Gemmill A, Moller A-B, Fat DM, et al. A Bayesian approach to the global estimation of maternal mortality. The Annals of Applied Statistics. 2017 Sep;11(3):1245–74. Available from: http://dx.doi.org/10.1214/16-aoas1014

29. Hill K, You D, Inoue M, Oestergaard M. Technical Advisory Group of United Nations Inter-agency Group for Child Mortality Estimation. Child mortality estimation: accelerated progress in reducing global child mortality, 1990–2010. PLoS Medicine. 2012 Aug 28;9(8):e1001303. Available from: http://dx.doi.org/10.1371/journal.pmed.1001303

30. United Nations Inter-agency Group for Child Mortality Estimation. Child mortality estimates. 2017 [cited 2018 22nd January]. Available from: http://www.childmortality.org/index.php?r=site/index

31. Paciorek CJ, Stevens GA, Finucane MM. Children's height and weight in rural and urban populations in low-income and middle-income countries: a systematic analysis of population-representative data. Lancet Global Health. 2013. Available from http://dx.doi.org/10.1016/S2214-109X(13)70109-8

32. Morris SS, Black RE, Tomaskovic L. Predicting the distribution of under-five deaths by cause in countries without adequate vital registration systems. International Journal of Epidemiology. 2003 Dec;32(6):1041–51. Available from: http://dx.doi.org/10.1093/ije/dyg241

33. Global Burden of Disease (GBD) 2015 Tobacco Collaborators. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990–2015: a systematic analysis from the Global Burden of Disease Study 2015. 2017 [cited 2018 25th January]. Available from: https://doi.org/10.1016/S0140-6736(17)30819-X

34. Gething PW, Patil AP, Smith DL, Guerra CA, Elyazar IR, Johnston GL, et al. A new world malaria map: Plasmodium falciparum endemicity in 2010. Malaria Journal. 2011;10(1):378. Available from: http://dx.doi.org/10.1186/1475-2875-10-378

35. Shmueli G. To explain or to predict? Statistical Science. 2010 Aug;25(3):289–310. Available from: http://dx.doi.org/10.1214/10-sts330

36. Browne MW. Cross-validation methods. Journal of Mathematical Psychology. 2000 Mar;44(1):108–32. Available from: http://dx.doi.org/10.1006/jmps.1999.1279

37. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Statistics Surveys. 2010;4(0):40–79. Available from: http://dx.doi.org/10.1214/09-ss054

38. Foreman KJ, Lozano R, Lopez AD, Murray CJ. Modeling causes of death: an integrated approach using CODEm. Population Health Metrics. 2012 Jan 6;10(1). Available from: http://dx.doi.org/10.1186/1478-7954-10-1

39. Barai S, Reich Y. Ensemble modelling or selecting the best model: Many could be better than one. AI EDAM. 1999 Nov;13(05). Available from: http://dx.doi.org/10.1017/s0890060499135029

40. Hsieh W, Tang B. Applying neural network models to prediction and data analysis in meteorology and oceanography. Bulletin of the American Meteorological Society. 1998;79(9):1855–70. Available from: https://doi.org/10.1175/1520-0477(1998)079<1855:ANNMTP>2.0.CO;2

41. Gather Working Group. Guidelines for accurate and transparent health estimates reporting: GATHER explanation and elaboration. 2016 [cited 2018 25th January]. Available from: http://gather-statement.org/

42. Stevens GA, Alkema L, Black RE, Boerma JT, Collins GS, Ezzati M, et al. Guidelines for accurate and transparent health estimates reporting: the GATHER statement. PLOS Medicine. 2016 Jun 28;13(6):e1002056. Available from: http://dx.doi.org/10.1371/journal.pmed.10020

43. Rudan I, Chan KY. Global health metrics needs collaboration and competition. Lancet. 2015 Jan;385(9963):92–4. Available from: http://dx.doi.org/10.1016/s0140-6736(14)62006-7

44. Kovacs SD, Mullholland K, Bosch J, Campbell H, Forouzanfar MH, Khalil I, et al. Deconstructing the differences: a comparison of GBD 2010 and CHERG's approach to estimating the mortality burden of diarrhea, pneumonia, and their etiologies. BioMed Central Infectious Diseases. 2015 Jan 16;15(1). Available from: http://dx.doi.org/10.1186/s12879-014-0728-4

45. Mola G, Kirby B. Discrepancies between national maternal mortality data and international estimates: the experience of Papua New Guinea. Reproductive Health Matters. 2013 Jan;21(42):191–202. Available from: http://dx.doi.org/10.1016/s0968-8080(13)42725-8

46. Mathers C, Stevens G, Hogan D, Mahanani WR, Ho J. Global and regional causes of death: patterns and trends, 2000–15. In Jamison DT, Gelband H, Horton S, Jha P, Laxminarayan R, Mock CN, Nugent R, editors. 2018. Disease control priorities: improving health and reducing poverty. Disease Control Priorities, Third ed. (Volume 9). Washington, USA: World Bank; 2017:69–104.

47. GBD 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017;390(10100): 1151–210. Available from: http://dx.doi.org/10.1016/S0140-6736(17)32152-9

48. World Health Organization. The utility of estimates for health monitoring and decision-making: global, regional and country perspectives. Report of a Technical meeting, 2015. | WHO, Glion sur Montreux, Switzerland, 24–25 June 2015 [cited 2018 5th November]. Available from: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=2ahUKEwjJ9qDqvcreAhVsy1QKHSQzDvoQFjABegQIARAC&url=http%3A%2F%2Fapps.who.int%2Firis%2Fbitstream%2F10665%2F182157%2F1%2FWHO_HIS_HSI_2015.6_eng.pdf&usg=AOvVaw2i9J6Ov-PQQVoZltYFkQEx

49. United Nations Statistics Division. Revised list of Global Sustainable Development Goal indicators. 2017 [cited 2018 5th November]. Available from: https://unstats.un.org/sdgs/indicators/Official%20Revised%20List%20of%20global%20SDG%20indicators.pdf