Committee for the Coordination of Statistical Activities

Conference on Data Quality for International Organizations
Rome, Italy, 7 and 8 July 2008

Session 4: Best practices in imputation and estimation

WHO practices in imputation and estimation

Carla Abou-Zahr, WHO

The production and dissemination of health statistics for health action at the country, regional and global levels are core WHO activities mandated by the Member States in the Constitution. The Organization's strategy in health statistics focuses on strengthening country, regional and global health information systems. In doing so, WHO positions itself as a partner in the international statistical system and its adherence to the principles governing international statistical activities.[1]

Data imputation and estimation are but one step in the production of valid, reliable and comparable statistics. The quality of imputation and estimation processes depends critically on two previous steps, namely, primary data collection using standardized definitions and methodologies, and data compilation. A final step involves the dissemination and communication of health statistics to interested partners, in appropriate formats. Thus, WHO's practices in health statistics can be summarized around four areas:

- Data generation;
- Data compilation;
- Data synthesis, analysis, imputation and estimation;
- Information communication and dissemination.

This paper summarizes WHO activities across all four areas, with a focus on the strategies used to assure quality at each stage. However, the primary focus of the paper is on step three, namely, data synthesis, analysis, imputation and estimation.

Data generation

Norms and standards for data collection

The generation of data that are comparable over time and within and across countries is dependent on the use of standard nomenclatures, terminologies, definitions and data collection tools. WHO has a key role in setting norms and standards for health-related data generation in countries, covering all diseases and health programmes. Major resources are invested in the

[1] http://unstats.un.org/unsd/methods/statorg/Principles_stat_activities/principles_stat_activities.pdf

development of standards for indicators and associated data collection methods.[2] The single most important standardization mechanism is the International Classification of Diseases and Related Health Problems (ICD) which is the international standard diagnostic classification for all general epidemiological and many health management purposes. These include the analysis of the general health situation of population groups and monitoring of the incidence and prevalence of diseases and other health problems in relation to other variables such as the characteristics and circumstances of the individuals affected. The ICD is used to classify diseases and other health problems recorded on many types of health and vital records including death certificates and hospital records. In addition to enabling the storage and retrieval of diagnostic information for clinical and epidemiological purposes, these records also provide the basis for the compilation of national mortality and morbidity statistics by WHO Member States.

In setting health information standards WHO collaborates with partners including WHO collaborating centres, UN agencies, and academic institutions. Standards are regularly updated to ensure relevance to changing epidemiological conditions and are subject to peer review by multidisciplinary experts from around the world.

Primary data collection

WHO provides technical and financial support to primary data collection. Examples include international survey programmes (World Health Survey, Global Youth Tobacco Survey, Global School Health Surveys), multi-country studies on specific topics (ageing and health, tobacco use, child and reproductive health), surveillance of risk factors for noncommunicable diseases (STEPS)), and reporting based on health services and disease surveillance. However, WHO relies heavily on data generated through household surveys supported by USAID (Demographic and Health Surveys) and UNICEF (MICS surveys). An important contribution of WHO is to ensure that data are collected on the full array of health challenges that countries face in an era of rapid health transition, in which acute infectious and chronic noncommunicable conditions coexist.

In order to generate health statistics covering the full range of issues including mortality, morbidity, use of health care services, risk factors and financing, it is necessary to draw from a range of data sources, including censuses, surveys, civil registration, individual health status records, records from health facilities and administrative records. WHO aims to promote an optimal balance in the use of different data sources, in line with the Health Metrics Network framework,[3] which, developed in close collaboration with WHO, aims to guide the development of health information systems in low-income and middle-income countries.

Data compilation and sharing

Databases

[2] For instance, WHO has developed standards for the generation of data for service provision (e.g. the Expanded Programme on Immunization, the TB treatment monitoring system based on health facility records), the incidence or prevalence of diseases (e.g. HIV/AIDS surveillance, integrated disease surveillance and response), risk factor prevalence (e.g. for noncommunicable diseases), and overall health status measurement (e.g. through population-based surveys).

[3] http://www.who.int/healthmetrics/documents/hmn_framework200609.pdf

The WHO Statistical Information System (WHOSIS) brings together core health statistics for the 193 WHO Member States and comprises more than 70 indicators, including mortality, morbidity, burden of disease, coverage of health care, risk factors, inequities in health, and health care financing. In addition, WHO programmes maintain a large number of disease-specific databases comprising many more detailed indicators. Currently, the Organization is engaged in a systematic review of data compilation activities across the Organization and has identified the need to increase standardization and interoperability so as to enhance data sharing and exchange both internally and externally with databases maintained by other agencies. In some cases, databases are jointly maintained with partners. Examples are the child and maternal health and mortality databases maintained by UNICEF in collaboration with WHO, the World Bank and the UN Population Division; and the joint UNICEF/WHO water and sanitation monitoring database and the WHO/UNICEF immunization coverage database.

Quality ascertainment

WHO is working to ensure that the data produced by the Organization meet quality standards of reliability, transparency and completeness using the Data Quality Assessment Framework (DQAF) as a guide. The Organization is also collaborating with the World Bank to support countries in improving the quality of health statistics using the General Data Dissemination Framework.

Data-sharing and transparency

WHO promotes full transparency of data used to generate health statistics and adherence to basic principles and criteria for data quality set by statistical constituencies. An important aspect is data-sharing: underlying data sets used for the production of health statistics should be in the public domain, well documented with meta data, and presented in a standardized manner. This should include individual or household-level microdata from household surveys, and data reported by facilities and used by ministries of health to produce health statistics.

WHO is creating a Global Health Observatory to enhance access to – and analyses of – all WHO's information on global health situations and trends. The Observatory will be an information resource for all those working on health issues, combining easy access to health data through one web portal, with a series of analytical reports. Building upon existing health information activities it will provide a mechanism to improve communication of statistics, enhance data quality and international comparability, strengthen collaboration with partners, including UN agencies, private sector and academic institutions, and identify data gaps in countries and promote ways to fill them.

Data analysis, synthesis, imputation, and estimation

Production of comparative statistics

Depending on the nature of the data and available data collection strategies, health statistics can be divided into three types: unadjusted, adjusted, and predicted. *Unadjusted* health statistics are derived directly from primary data collection with no adjustments or corrections. *Adjusted* health statistics comprise indicators that have been subject to some kind of analytic technique. Adjustments may involve corrections to deal with known biases. Adjustments may also relate to the issue of indirect techniques to estimate quantities of interest such as the use of data on

children ever born and children still alive to estimate levels of child mortality. *Predicted health statistics* (also known as imputed statistics) are based on a model relating the quantity of interest to covariates. There are two widely used types of predicted statistics; one is forecasting, using past relationships to predict into the future; another type of prediction is used to generate figures for indicators in settings without adequate primary data (sometimes called "farcasting").[4]

For some health issues, it is possible to compile comparable health statistics that are based solely on statistics reported by countries without further corrections or predictions – unadjusted statistics. This is most likely to happen for statistics from high-income countries (e.g. density of health workers) or for notifiable diseases, e.g. polio. However, the use of unadjusted statistics is generally the exception. Most often, the production of a health statistic requires some kind of adjustment in order to enhance comparability over time, within parts of a country, or between countries. Adjustments may involve the reconciliation of data from different sources, correction such as age standardization, methods to deal with non-representativeness of samples. Prediction or imputation is used to fill data gaps and to produce summary measures of population health such as healthy life expectancy (HALE). It is evident that the quality of prediction is highly dependent on the quality of the input data available: better empirical data support better modelling efforts and vice versa. In practice, reporting of the health-related MDGs relies heavily on predicted statistics, both because of the need to produce an estimate for more recent periods than covered by the data collection method (estimates of child mortality derived from household surveys, for example, actually relate to a period of time generally around three years prior to the survey), or because of the need to impute missing values (for example, for maternal mortality in many countries).

Many WHO programmes develop and use methods and tools for imputation and estimation. Some headquarters programmes have relatively strong in-house expertise (e.g. TB), some have strong partnerships with other agencies (e.g. AIDS), and some rely heavily on assistance from the WHO Department of Health Statistics and Informatics (HSI). In general, there is strong involvement of academic institutions. Annex A provides a summary of the main expert groups involved in the production of corrected and predicted health statistics.

Quality assurance and country consultation

To ensure the quality of estimates produced and used by WHO a four-step quality assurance framework is in place at headquarters. The standard procedure for the estimation of disease burden and risk factors, including the coverage of interventions, should include:[5]

- a high-quality accessible database;
- an independent review group;
- published methods of estimation;
- internal WHO clearance by the designated unit, generally HSI .

Country consultation is the final step before the public release of country-level estimates. This is an iterative process between Member States, WHO country and regional offices and headquarters. It serves multiple purposes, including obtaining approval from Member States, providing

[4] Murray, CJL (2007) Towards good practice for health statistics: lessons from the Millennium Development Goal health indicators ***Lancet:*** 369: 862–73,

[5] *WHO Information Note* 42/2005.

feedback on the quality of data collection and reporting by countries, and strengthening country capacity to reproduce, produce and use estimates. All methods and assumptions are, to the extent possible, provided in a user-friendly format. Critically, the country consultation process should be accompanied by capacity-building. In some disease areas extensive investment have been made in software tools and training workshops of country staff.

The results of the imputation process may differ from a country's own officially reported statistics. While the primary goal is to reach consensus on the basis of a mutual understanding of the methods used, regional offices sometimes publish both reported or registered health statistics and WHO estimates (e.g. for life expectancy). Headquarters has often included disclaimers alerting readers to possible differences.

External advisory bodies

In addition to the programme-specific quality assurance mechanisms described above and in Annex A, the WHO Director-General has established an independent Advisory Committee on Health Monitoring and Statistics (ACHMS). The ACHMS provides inputs to WHO on its work to monitor the health situation and trends, including Global Burden of Disease, Global Health Observatory, and World Health Statistics. The group also provides the WHO Director-General with advice on scientific and technical issues related to health statistics, including overall strategy and procedures of the Organization, and current and future issues related to estimation of disease burden and risk factors.

The ACHMS has expressed concern about the poor availability and quality of data and stressed the importance of mortality and cause-of-death statistics, disaggregated by age, sex, cause, and socioeconomic status. The ACHMS made several recommendations in this regard:

- expanded use of the WHO standardized verbal autopsy tools and continued efforts to improve such tools through methodological work;
- support to efforts by the Health Metrics Network and other partners to strengthen monitoring of mortality in countries with limited information using methods appropriate for the specific country situation;
- development of specific goals and targets to improve the development and use of valid methods to ascertain cause-of-death information;
- an expansion of the WHO mortality database, including data sources other than vital registration systems, such as those from population censuses, demographic surveillance sites, and sample registration systems.

Communication of information

WHO is the world's main source of comparable health statistics. Disease programmes invest in regular annual publications, e.g. the TB annual report, or in special reports. Each regional office publishes a statistical summary every year before the regional commission meeting in a similar format for all regions. In 2005, WHO regional offices and headquarters agreed on a set of core health indicators resulting in the annual publication of the World Health Statistics. Statistical annexes are also included in the World Health Report and in the regional situation analyses. Many programs produce country health profiles. The main challenge is to ensure consistency and sustain regular updating. It is also critical to empower users by enhancing efforts to communicate the underlying data availability and quality, including by producing uncertainty ranges and the use of explanatory texts in publications such as the World Health Statistics report. World Health

Statistics 2008 included, for the first time, ten highlights in global health than covered not only the current situation and trends on important health issues but also an explanation of the underlying data sources and estimation methods.

The Global Health Observatory will enhance WHO's dissemination and communication of health statistics in two ways:

- a web portal providing one entry-point to WHO's health statistics and analyses
- a series of reports.

The web portal will provide the world with up-to-date, easily accessible information on priority health topics. This portal provides public access to an interactive database of all health data and statistics from across WHO. It will also be a rich source of information and evidence on diseases, interventions and progress towards major health goals. Up-to-date country health profiles − compilations of health statistics that cover all major health areas, from acute disease outbreaks to chronic conditions - will be available in both standardized format (e.g. for MDGs, health systems, burden of disease) and dynamic formats, in which users can be define the contents from a range of indicators and determinants.

Annex A
Overview of interagency monitoring/estimation groups – purpose and membership

Subject area	Expert Groups	Purpose	Membership
Child health	Interagency Coordination Group on Child Mortality Estimation (IGME) with Technical Advisory Group (TAG)	● Develop joint estimates for under-five, infant and neonatal mortality	International organizations (IGME) (UNICEF, WHO, The World Bank, UN Population Division); academia and institutions (TAG) (Harvard, LSHTM etc.)
	Child Health Epidemiology Reference Group (CHERG)	● Develop cause-specific mortality estimates ● Review evidence of the impact of interventions on child survival	International organizations (UNICEF, WHO, CDC and others); academia and institutions (JHU, LSHTM, others)
	Child Survival Countdown – Technical Working Groups	● Report on progress in child survival intervention coverage indicators ● Review technical issues related to key child survival indicators	International organizations (UNICEF, WHO, and others); academia and institutions (various)
	Interagency group for immunization estimation	● Develop joint immunization estimates	International organizations (UNICEF, WHO)
	Quantitative Immunization and Vaccine related Research (QUIVER)	● Oversee the development and improvement of methods and data to evaluate the burden of vaccine-preventable diseases ● Standards for economic evaluation ● Tools for country decision-making	WHO and 12 members from academia and other institutions
Maternal mortality	Interagency group for maternal mortality estimation and trend analysis	● Develop joint maternal mortality estimates and new methodology for trend analysis ● Prepare regional workshops to explain methodology and promote data analysis and use	International organizations (UNICEF, WHO, UNFPA, World Bank, UN Population Division); academia and institutions (Harvard and others)
Malaria	Roll Back Malaria - Malaria Monitoring and Evaluation Reference Group (MERG) WHO Strategic and Technical Advisory Group for Malaria (STAG)	● Develop standard indicators, methodologies and monitoring tools ● Develop joint malaria coverage and burden estimates ● Produce joint biennial reports (e.g. World Malaria Report) ● Prepare regional workshops to promote data analysis and use ● RBM work based on activities of 6 task forces:	International organizations (UNICEF, WHO, USAID, The World Bank, The Global Fund, MACEPA) Academia and institutions (CDC, LSHTM, ORC Macro, Harvard, others)
Neglected Tropical Diseases	Scientific and Technical Advisory Group (NTD - STAG)	● To estimate the burden of disease work part of its mandate	International organizations, academia, other institutions; Includes advocacy elements
	Leptospirosis Epidemiology Reference Group (LERG)	Develop estimates of burden of leptospirosis	Technical experts in disease modeling, zoonotic diseases, burden estimation
AIDS	UNAIDS Reference Group on Estimates, Epidemiology and Modelling	● Develop methods and tools to estimate prevalence, incidence and mortality ● Review UNAIDS and WHO on estimates	International organizations (UNAIDS, WHO, UNICEF) Academia and institutions (secretariat Imperial College London, East West Centre, LSHTM, CDC,

		● Facilitate country capacity building in surveillance data quality assessment, analysis and estimation work	Futures and others)
	HIV/AIDS Monitoring and Evaluation Reference Group (MERG)	● Develop standard indicators, methodologies and monitoring tools ● Approval of new indicators and guides ● Evaluation of programmes	International organizations (UNAIDS, UNICEF, WHO and others) Academia and institutions (various)
Risk factors	Joint Monitoring Program (JMP) for Water Supply and Sanitation and Technical Advisory Group	● Develop joint water/sanitation estimates and methodology for trend analysis ● Technical Advisory Group advises on development of new methodologies, indicators and monitoring tools ● Prepare regional workshops to explain methodology and to support data analysis and use	International organizations (UNICEF, WHO, World Bank, USAID) Academia and institutions (LSHTM and others)
	Food-borne disease epidemiology reference group (FERG)	● Estimate the global burden due to foodborne diseases	International organizations, academia and other institutions
MDG	Interagency Group on MDG indicators	● Reach consensus on indicators, data, metadata and reporting	UN agencies, led by UN Statistical Division