Contributors and sources: The authors are clinical epidemiologists who are involved in critical review of medical and surgical studies as part of their educational and editorial activities. In that context, they have collected some of the examples presented here. Other examples were collected by searching titles of records in PubMed using the textwords “misleading,” “biased,” and “spin” using the Related Articles feature for every pertinent hit, and by noting relevant references from retrieved papers. VMM and GHG wrote the first draft of the article. All authors contributed to the ideas represented in the article, made critical contributions and revisions to the first draft, and approved the final version. GHG is guarantor.

Funding: VMM is a Mayo Foundation scholar. PJD is supported by a Canadian Institutes of Health Research senior research fellowship award. MB was supported by a Dettwiler fellowship, Royal College of Physicians and Surgeons of Canada and is currently supported by a Canada Research Chair.

Competing interests: VMM, RJ, HJS, PJD, and GHG are associate editors of the ACP Journal Club and Evidence-Based Medicine. JB and RJ edit an evidence based medicine journal in Poland. MB is editor of the evidence based orthopaedic trauma section in the Journal of Orthopaedic Trauma.

7 Clarke M, Alderson P. Clarke’s Education and debate section in reports of controlled trials published in general medical journals. JAMA 2002;287:2799-801.

(accepted 15 September 2004)

Monitoring global health: time for new solutions

Christopher J L Murray, Alan D Lopez, Suwit Whipulphprasert

Improved global health monitoring requires new technologies and methods, strengthened national capacity, norms and standards, and gold standard global reporting. The World Health Organization’s many functions limit its capacity for global reporting, and a new global health monitoring organisation is needed to provide independent gold standard health information to the world.

Sound information on financial and human resources invested in health, health interventions delivered to people in need, and the impact of these efforts on people’s health is critical for planning health systems, implementing programmes, epidemiic response, allocating budgets for research and development, monitoring progress, and evaluating what works and what does not. Although all countries collect a wide range of health information through registers, surveys, and vital registration systems, huge gaps hinder our ability to respond to global health challenges, which are alarming at a time when global investments in AIDS, tuberculosis, and malaria are increasing and when there is renewed focus on health goals as exemplified by the UN Millennium Declaration.1 The availability of valid, reliable, and comparable health information to inform local, regional, national, and global decisions can be furthered through four interconnected efforts: improving the technology and methods for population health measurement; strengthening national capacity and motivating governments to collect and analyse useful health data; establishing global norms and standards for what is the core health related measurements and how to measure them; and reporting to the globe valid, reliable, and comparable assessments of inputs, service delivery, and achievements for health. Although many challenges and initiatives are under way for the first three of these components, the fourth area is currently the weakest and getting worse, not better. We explore the problem and the potential solution to global monitoring and evaluation and briefly review the other three areas as necessary.

Technology and methods of health measurement

For several critical measures of health, delivery of interventions, and financial and human resources including many millennium development goal indicators, the current measurement technology is inadequate. Although validated methods have been developed to measure child mortality through household surveys in settings where vital registration is incomplete or non-existent,14 no adequate method...
exists to measure adult mortality in such settings. Although antibody tests for HIV mean that the prevalence of infection in the population can be ascertained from sample surveys, no affordable and feasible methods are currently available to assess tuberculosis in a community. Advances in immunology, proteomics, genomics, metabolomics, survey science, and statistical methods may result in new technologies or methods coming on line in the next decade that will dramatically improve our ability to monitor population health.

**Strengthening capacity**

Strengthening national capacity to collect and analyse data is essential. Intense efforts have been made to strengthen data collection for particular vertical disease programmes, such as polio eradication or selected sentinel communities, but progress in national collection of health data and capacity for analysis in low income and many middle income countries has been slow. One exception is the development of the Thai Health Foundation, where a 2% excise tax on tobacco and alcohol funds a major programme for improvement in the national health information system. Investments and efforts at technical assistance to build national health information capacity have continued over the past 30-40 years, with only limited success. Efforts to strengthen national capacity may be invigorated by the Health Metrics Network (www.who.int/healthmetrics), which seeks to catalyse the development of health information systems in developing countries. In current plans, the network will focus on capacity building in five to seven countries per year. The World Health Organization along with other partners has a lead role in providing technical assistance and guidance in developing national capacity for health information. Progress is often hindered by the difficulty of showing to national decision makers that good health data can strongly support decision making. In our experience, enhanced global reporting will increase government commitment to collect high quality data.

**Norms and standards**

A third building block for effective health information is the establishment of global norms and standards on key indicators for different health programmes and for health systems overall, the best measurement methods for these indicators given current technology and analytical methods, and standardised definitions and classification systems. As illustrated by the International Classification of Diseases and Injuries over the past 50 years, WHO can play a powerful part in all this. For overall health statistics and many disease specific or risk factor specific areas, WHO can and should remain the leading institution. Nevertheless, WHO has shifted away from work on global norms and standards to emphasise country implementation of disease programmes, which will create a void that other institutions may need to fill.

**Global reporting**

The fourth component—critical if health information is to have an impact locally, nationally, or globally—is the creation and dissemination of gold standard information on key indicators of inputs, achievements, and impacts of health interventions. Reporting information that is valid and reliable and can be benchmarked meaningfully is essential for monitoring progress and evaluating what works and what does not. Unless health information is disseminated to the public, the scientific and public health community, and decision makers through multiple channels, including the media, scientific journals, and other documents, it more often than not remains unused in statistical abstracts or spreadsheets in health ministries.

The sum of deaths claimed by different WHO programmes exceeded the total number of deaths in the world

WHO is the major actor in global health reporting because it is the leading agency in the United Nations system working on health and because in many cases no credible alternative exists. In some cases, WHO undertakes this reporting function in partnership with others such as UNICEF or UNAIDS, but at present the default expectation is that WHO should report gold standard information to the world. Because of this expectation, we focus much of our analysis of global health reporting on the structural challenges that WHO faces in this role before proposing potential solutions.

Over time and across technical areas, WHO’s performance in global reporting has varied tremendously. For example, for child and adult mortality, WHO has made systematic efforts to collate and analyse all available data sets and has published abridged life tables for all countries. In contrast, for one of the millennium development goal indicators, the prevalence of malaria, WHO simply reports country statistics irrespective of a wide range of known biases. As a consequence, Nigeria has a rate of 30 cases per 100 000 per year and Guatemala has 386/100 000. At a fundamental level, the architecture of global institutions has an essential flaw with regards to monitoring and evaluation, which must be addressed. We identify the core reasons for this problem and offer potential solutions. Addressing the problems with global reporting will, we believe, also fuel greater commitment among countries to strengthening national health information systems.

Too many roles for WHO

WHO very often finds itself in the multiple roles of global advocate, provider of technical assistance to countries, monitor of progress towards targets, and evaluator of what works and what does not. The commitment and dedication of WHO’s staff working in the technical departments is unquestionable. The problem is that staff representatives inevitably feel the tension between advocacy, monitoring, and evaluation. In other arenas, such as business, it would
be unthinkable to ask a company to audit themselves. Two examples can serve to illustrate the consequences of this problem.

In the early 1990s, the global burden of disease study was initiated in part because the sum of deaths claimed by different WHO programmes exceeded the total number of deaths in the world several times. The intense pressure on technical programmes to keep their figures as large as possible was evident. Crudely, size of problem often translated into dollars from donor agencies. The global burden of disease study, published in 1996, set new benchmarks for internal consistency, comparability, and comprehensiveness of epidemiological information.

Since 1998, WHO has adopted and implemented the study's approach to producing coherent epidemiological information. Given the importance of figures for advocacy, technical programmes would exert intense competitive pressure on the epidemiology and burden of disease team that was charged with bringing together WHO's annual assessments. These pressures were withstood because of the strong commitment of the senior management to valid, reliable, and comparable epidemiological information. This process, although imperfect, meant that internally consistent, comparable, and comprehensive information on incidence, prevalence, mortality, and disabling sequelae by age and sex for 14 epidemiological subregions of the world were published each year. Over the past year, as the focus of WHO has shifted to technical assistance for countries, the epidemiology and burden of disease team has been reduced from 22 to two staff members, and it seems unlikely that the effort to produce coherent comparable epidemiological data by major causes of death and ill health will continue.

WHO's tuberculosis programme provides another example of the tensions that inevitably exist in an organisation simultaneously developing advocacy material, providing countries with technical assistance, monitoring progress towards the global targets, and evaluating its own directly observed treatment, short course (DOTS) strategy. Direct measurements of tuberculosis in populations come from four potential sources: registered tuberculosis deaths, notified new cases of tuberculosis, purified protein derivative standard (PPD) skin testing of BCG scar negative children, and sputum prevalence surveys. For most low income countries, the only source is case notifications. WHO has estimated true incident cases from case notifications if rates over 100% are accepted without scrutiny?

The figure shows, by using isoquants, all possible combinations of true incidence and case detection rates for each year. As estimates for a given year are arbitrary, estimates of trends in case detection rates over time have no empirical basis—the isoquants in the figure are consistent with increasing or decreasing trends in case detection rates, and no information is available on the true trend. It should also be noted that WHO's assumed case detection rates and true incidence for the same year differ in the published global tuberculosis reports and on its website. Serial guessing is not a sound basis for monitoring progress towards a global target of 70% case detection. Because the WHO programme is also the global advocate for tuberculosis control, it cannot and does not bring attention to the fact that essentially no empirical basis exists to assess the trend in case detection in regions where tuberculosis is most prevalent, including sub-Saharan Africa.

National politics and impossible figures

WHO is a voluntary association of sovereign states; its owners are the 192 member states. More practically, strategy is set by the ministers of health or their representatives, who attend the executive board twice a year, and to the World Health Assembly in May of each year. Irrespective of legal mandates, if a powerful country disputes country specific figures produced by WHO, it can bring strong pressure to bear on the organisation to change the data. Everyone who has worked in a senior position in the United Nations recognises this problem.

A standard compromise used in the UN, including WHO, is to report regional figures that have been analysed and corrected for known biases and national figures that are simply the figures sent by member states. This situation produces many examples of bizarre results. Regional totals of disease incidence or patients receiving treatment often do not equal the sum of published country specific figures. Another example is the publication of impossible figures, such as case detection rates for smear positive tuberculosis greater than 100% for Oman, Chile, Honduras, and Algeria, among others. How can we believe any of the national figures published for case detection if rates over 100% are accepted without scrutiny?
As novel institutions such as the Global Alliance for Vaccines and Immunizations and the Global Fund for AIDS, Tuberculosis and Malaria propose to link disbursements to achievements in delivering interventions, the pressure at the national level to provide biased data will intensify. Even before the global alliance, no relation existed between reported changes in immunisation coverage and changes as measured through household surveys. The global alliance recognises the potential problem of providing performance related disbursements and asking recipients if they have achieved their targets. Its data quality audits show profound problems and opportunities for distortion. Still, vaccine coverage has not been accessed independently. The same problems will apply to investments of the global fund.

Pressure from countries on WHO's leadership means a constant possibility that monitoring and evaluation can be distorted. The experience during the severe acute respiratory syndrome (SARS), when strong WHO leadership withstood pressure from both China and Canada and issued travel warnings, shows that courageous leadership can withstand political pressure. The experience of SARS and outbreaks more generally points to two different dimensions of the problem.

Firstly, WHO may be able to resist political pressures for brief periods of time, but this becomes increasingly difficult as the time span gets longer. Because many reports of outbreaks are picked up from local media reports and not national governments, and because of the intense media focus on outbreaks, the time available for political pressure on WHO leadership is much shorter. It seems likely that WHO can, especially with strengthened international health regulations, fulfill the need for reporting global outbreaks. However, for monitoring inputs to health systems, delivery of health interventions for a wide range of programmes, and progress on overall health, the dynamics are different. As the time frame is longer, the potential for manipulating the data is much greater.

Secondly, structural problems of monitoring and evaluation for organisations such as WHO can be overcome temporarily or at least attenuated by strong leadership. But we cannot expect always to have organisational leadership that is willing to take on this challenge. We should recognise the important roles of WHO in building national capacity for health information, establishing norms and standards for measurement, and setting agendas for research and development on new technologies and methods. At the same time, we have to recognise that WHO is ill suited for the role of global monitoring and evaluation of health.

Potential solutions

To sustain increased investments in global health, gold standard information is essential. In the long run, such information will require better measurement methods and technologies, capacity strengthening in developing countries, and global norms and standards. To fulfill our need for gold standard information in the short term and to fuel government commitment to better health information in the long term, the institutional problems of global reporting by WHO and other UN agencies need to be solved. Although our analysis has stressed the structural issues of WHO, the problems of mixing advocacy, technical assistance, monitoring and evaluation roles, and maintaining independence from country pressure apply to other agencies and organisations as well.

Solutions such as creating an independent arm of the WHO that reports directly to the executive board, a strategy adopted by the World Bank and the International Monetary Fund, are possible but pose nearly insurmountable political challenges. Initiatives that are specific to a disease or intervention to undertake independent monitoring to a gold standard may be helpful but will miss out on the many opportunities for improved global reporting that are common across programmes. We believe that the only viable solution will be to create a new, independent, health monitoring organisation.

The objective of this body would be to report regularly to the world on what is spent on health, what health services are delivered, and the impact of these efforts on population health. This organisation would be small as its main role would be to collate, analyse, and disseminate the best available evidence. Much of this work would be in close partnership with various actors such as the WHO technical programmes, the Global Alliance for Vaccines and Immunizations, and the Global Fund for AIDS, Tuberculosis and Malaria. To be effective, the organisation would need to be as sheltered as possible from the needs of advocacy on the one hand and country political interference on the other.

Those familiar with the complex governance issues that new entities such as the global alliance or the global fund have faced will recognise that solving the governance and financing issues for this organisation will not be easy, but it can be done. Success of such an organisation would depend on several key factors. Firstly, all representatives from the key stakeholders in global reporting would need to have a voice in the governance of this effort. Key stakeholders would include national governments, multilateral institutions (WHO, UNICEF, UNAIDS, UN Development Programme, World Bank, European Union, and others), bilateral donor agencies, a range of non-governmental organisations, and the research community.

Secondly, to be effective such a health monitoring organisation would have to be committed to the principles of validity, reliability, comparability of figures, an
A new global health monitoring organisation is needed to provide independent gold standard health information to the world.

Improved global health monitoring requires new technologies and methods, strengthened national capacity, norms and standards, and gold standard global reporting.

WHO’s ability to undertake independent global reporting is limited by its simultaneous roles as global advocate, technical adviser to countries, monitor, and evaluator—and by intense political pressure.

Progress could undermine the consensus that investing in global health is vital. It is in the interest of all the global community to invest in solid independent monitoring and reporting.

Acknowledgements: We thank Julio Frenk, Sudhir Anand, Sue Goldie, and Anthony Rodgers for helpful comments. Jesse Abbot Klafter provided invaluable assistance on the analysis of tuberculosis coverage.

Competing interests: None declared.

Summary points

- Improved global health monitoring requires new technologies and methods, strengthened national capacity, norms and standards, and gold standard global reporting.
- WHO’s ability to undertake independent global reporting is limited by its simultaneous roles as global advocate, technical adviser to countries, monitor, and evaluator—and by intense political pressure.
- A new global health monitoring organisation is needed to provide independent gold standard health information to the world.

Explicit data audit trail, and open consultation. As health information reaches a wider audience and touches on issues salient to everybody’s life, scrutiny of this information will intensify. The criticisms of the World Health Report 2000 and the subsequent recommendations of the scientific peer review group highlight the importance of total transparency in the process of measurement.

Validity and reliability are familiar concepts in health measurement. Comparability means that results should be reported in a manner that allows meaningful comparisons to be made between countries and over time. Committing to an explicit data audit trail is costly but essential. This means that every step in the development of a figure should be made publicly available, including the primary data, details on the methods used to analyse the primary data (including corrections for known biases), and adequate documentation of all steps in the analysis. The extraordinary commitment of the Human Genome Project to put all primary data in the public domain with effectively no time lag should be held up as a model to follow. Finally, open consultation means that both governments and the scientific community at large should be able to comment and critique figures that are published. Fostering healthy debate on measurements will lead to better data collection and analysis.

Thirdly, an independent monitoring organisation, although it could be a relatively small undertaking of the order of $30m-70m (£27m-38m; £39m-55m) per year, would require stable core resources. Securing the right combination of governance and stable core resources is the main challenge for creating such an organisation.

Without core resources, any organisation—regardless of governance structure—could be captured by its simultaneous roles as global advocate, technical adviser, and evaluator—and by intense political pressure. Fostering healthy debate on measurements will lead to better data collection and analysis.

In an era when the credibility of global health organisations is under attack, providing the public with credible, clear, and comparable health information will strengthen the commitment and resolution to scale up efforts on global health. At the Bangkok AIDS conference in July 2004, a journalist asked whether WHO figures on the delivery of antiretroviral therapy were “Enron-like.” Although such comments are clearly unfair, doubts about what has been achieved and whether we are pursuing the right strategies to make

leukaemia") were not predicted by animal studies using similar doses of vector.

One set of questions on toxicology related to gene transfer arises because most studies in humans—as with many other trials of hazardous agents—enrol participants with advanced illness. Such participants are likely to misinterpret the purpose of the trial as providing therapy rather than producing generalisable knowledge. Enrolment in studies on the safety of gene transfer is therefore susceptible to being based on “misinformed” consent. Also, participants who perceive a trial as providing therapy may be less willing to comply with intrusive procedures (for example, long term follow up and autopsy) that are aimed at testing safety. By policing consent procedures for language that promotes misconceptions about therapy, investigators may encourage participants to cooperate with a trial’s toxicological aspects. 

Premarketing studies of drugs often have insufficient power to expose rare adverse events; the collection of toxicity data is further hampered because gene transfer trials generally enrol participants with severe illness. For instance, attributing causes for adverse events is confounded by underlying medical conditions. Moreover, such populations are unlikely to survive and experience theoretically predicted latent adverse events. Therefore, many risks will only be characterised once gene transfer extends to populations with less severe medical conditions; patients and the public (rather than trial participants) will likely bear many of the risks involved in characterising latent toxicity.

Owing to the uncertainties and inexperience surrounding risks from gene transfer, systems may need to be established for postmarketing surveillance (for example, registries) and the long term follow up of trial participants. In the United States, such long term follow up is not mandatory, and anecdotal evidence indicates that it is not widely practised. In contrast, the United Kingdom and Australia (www7.health.gov.au/nhmrc/research/grap.htm) track the medical records of recipients of gene transfer. Follow up and postmarketing surveillance are potentially costly; can medicalise people’s lives, and infringe on their privacy. Nevertheless, spontaneous reporting of adverse events is unreliable for detecting latent adverse events, and more active measures may be necessary to protect the public, and patients and their descendants, should gene transfer extend to milder medical conditions.

Although recent trials confirm the feasibility of gene therapy, they also highlight that its risks are poorly understood. The task for researchers in gene transfer to humans is missing its first full stop (BMJ 2004;329:1441-4, 18-25 Dec). His correct address is k.fong@uc.ac.uk.

Monitoring global health: time for new solutions

The authors of this Education and Debate article (who argued that a new global health monitoring organisation is needed to replace the World Health Organization) would like to clarify for readers that they have all had recent links with WHO (BMJ 2004;329:1096-100, 9 Nov). Christopher J L Murray worked for WHO until 15 September 2003, Alan D Lopez worked for the organisation until 1 January 2002, and Suwit Wihulpopprasert has served on a number of advisory committees to WHO.

If the honey doesn’t get you, the bees will

A lapse in concentration by Harvey Marcovich, the author of the summaries on the BMJ Family Highlights page, led to the inadvertent omission of the word haemorrhagic in this summary on the BMJ Family Highlights page (BMJ 2004;329:1368, 11 Dec). The third sentence should have read:

"Computed tomography of the head and magnetic resonance imaging of the brain showed a large right temporo-occipital haemorrhagic infarct."

Cautious as teachers: the dissection room experience in Thailand

In this article by Andreas Winkelmann and Fritz H Gildner in our Christmas issue, we forget to carry out the authors’ wishes that we acknowledge Professor G H Schumacher from Rostock, Germany, as the provider of the photograph (BMJ 2004;329:1455-7, 18-25 Dec).