



Committee for the Coordination of Statistical Activities
Conference on Data Quality for International Organizations
Wiesbaden, Germany, 27 and 28 May 2004



Session 2: Coherence of data published by various international organizations on the same subject

Introducing a PARIS21 study of the international statistical system in relation to MDG indicators

“Monitoring the Millennium Development Goals: current weaknesses and possible improvements”

Neil Fantom, Development Data Group, World Bank

Background and objective

This study was commissioned by the PARIS21 Task Team on “Improving Statistical Support for Monitoring Development Goals”. The purpose was to provide a better understanding of processes behind the international statistical system, particularly as they relate to the key development indicators contained in the MDG monitoring set, identify weaknesses and propose low-cost solutions. The initial demand for this was discussed at the first International Roundtable on Development Results in June 2002 and Terms of Reference were subsequently agreed when the Task Team was formed in October 2002. Other work commissioned by the Task Team, which is convened by the World Bank, includes six country case studies and a study of internationally-sponsored household surveys.

This is a summary of the findings and recommendations of the study, which was conducted by Oxford Policy Management (OPM), consultants engaged by DFID. The process of compiling the study involved OPM obtaining detailed information from key Agencies with responsibility for specific indicators. Initial findings were discussed at a conference of international and national statistical agencies hosted by the World Bank in June 2003. Following this, refinements were made to the study to provide a broader and deeper set of information on the key conclusions were based. OPM produced their final report in January, and the results were used as a basis for some of the discussions at the Second Roundtable on Managing for Development Results, held in Marrakech in February 2004.

A briefing paper produced by OPM on the study is attached to this summary. The full report, together with detailed background data, is available from the PARIS21 web site, www.paris21.org.

Key findings of the study

1. In some cases there are inconsistent definitions applied by international agencies to the same indicator. The study documents examples where this arises because indicator definitions are not clear, not consistently applied by international organizations, or are not applied by countries when collecting data (HIV/AIDS indicators, literacy rates). It also examines indicator definitions that cannot always be readily applied to specific country circumstances, particularly where data are collected as a by-product of administrative systems (measles immunizations, births attended by skilled health personnel).

2. The use of different data sources to produce certain indicators makes cross-country comparability difficult. This is often dictated by data availability at country level; for example, welfare measures are based either on income or consumption, and there are often difficulties comparing the estimates calculated from surveys using each method. Differences between employment estimates derived from administrative records and labor force surveys are well known.

3. Since there is no systematic record of household surveys conducted by individual countries, systems which utilize estimates derived from household surveys occasionally fail to make use of all available data. Surveys with useful and valid data are sometimes missed by agency search mechanisms, even though sophisticated semi-formal sectoral networks exist, and agency field offices are often used to identify data sources. For example, the Pakistan Integrated Household Surveys (PIHS) were not identified as a useful source of mortality data by UNICEF and WHO, yet the PIHS is widely recognized to be one of the most important data sources for key indicators (including child and infant mortality) since 1996.

4. Although greater efforts are being made by agencies to more fully utilize data from household surveys, there are areas where the use of survey data to supplement data collected through administrative systems would improve the coverage and quality of data estimates. The main example chosen is school enrolment data, which could be supplemented or validated with attendance data from surveys (although it should be noted that UNESCO Institute of Statistics are already actively investigating ways to increase their use of survey data).

5. Collection of data using agency questionnaires is sometimes problematic. Although many agencies use their field offices, questionnaires do not always reach their intended recipient, or receive due attention when they do – particularly if they are long and onerous to complete. Countries do not always have adequate capacity to manage the questionnaire completion process well.

6. International population data are inconsistent, but are utilized heavily by some agencies, both to calculate indicator values for countries (e.g. as the denominator in six MDG indicators) and to generate regional and global estimates. There are three main sources: the UN Population Division, the World Bank, and the US Census Bureau. Data from these sources often differ; in one country in every six, for the year 2000, differences between the lowest and highest estimate vary by more than 10%. Differences in population estimates – particularly where the size of specific population sub-groups needs to be estimated – can make dramatic differences in indicator estimates.

7. Data presentation practices can cause misuse of indicator data. Estimates tend to be presented without full metadata detailing data sources, specific limitations, or freshness, and are sometimes presented representing a range of years. The result can be confusing to data users, who may not be able to, for example, easily distinguish between data updated last year on one web site, and data updated this year on another - and may even try to compare the two. The report documents cases where apparently different estimates for specific indicators and countries are reported by different agencies, but where the difference is actually caused by data freshness.

Some actions suggested in the report

1. Household surveys. Consolidate household survey networks, and create an international household survey database. Informal, semi-formal and formal networks already exist, and there are several examples of household survey databanks maintained by international organizations. The study proposes a joint effort to consolidate these initiatives, to improve data availability, more widely share knowledge and information, and to enable improved cross-country comparability. Proactive use of household survey data in improving estimates made primarily from administrative sources is also suggested.

2. Practices of the international statistical system. A systematic review of data collection and reporting practices. This should examine more closely the agency questionnaires and their use, and seek for a possible consolidation of reporting practices in order to minimize response burden and improve effectiveness. Rules and systems which define responsibilities of countries and agencies in the reporting process should also be reviewed to improve data quality.
3. Population estimates. Improvements in the use of international population estimates, including the provision of more complete documentation of sources and methods, and information on the precision of estimates. The study also suggests that it would be helpful for MDG reporting purposes if major producers of population data were to agree on a common set of estimates to be used in MDG calculations and reporting.
4. Methodology. Improvements in the management and use of common methodologies and definitions, including increased efforts to collate and promote definitions and guidelines (it is recognized that the UN MDG indicator metadata published in 2003 addresses some of these issues.) Greater attention needs to be paid to country-level data collection systems, to try to accommodate international definitions and requirements without distorting national requirements.
5. Dissemination. Increase the level of detail in metadata in published international databases, and make changes to data management and presentation practices to increase the information available to users and minimize the potential for misinterpretation and confusion, particularly when comparing data from different international sources.

For more information

Paris21 Task Team: <http://www.paris21.org>

Managing for Development Results: <http://www.mfdr.org>

Monitoring the millennium development goals

Current weaknesses and possible improvements – Briefing Note

January 2004



Introduction

This briefing note summarises the results of the international component of a study which was commissioned against a background of attempts by the Millennium Development Goals (MDG) Task Force, and others, with the aim of improving the quality of MDG indicators. The particular aim of the study was to provide a better understanding of the monitoring process and its standards, and to highlight areas for possible improvement.

Current weaknesses of the MDG indicators

The current weaknesses of the MDG indicators can be organised around two key questions:

1. what is the level of data available?, and
2. can a judgement be made about the quality of that indicator i.e. is the data comparable over time and between places?.

What has been the level of data availability?

During the 1990s the availability of data for the 48 MDG indicators saw improvement in almost all cases. However, there are very significant differences in availability between geographic areas and over time. These differences in availability and coverage are best explained by the “age” of the indicator, with indicators that have been widely used for many years having considerably better coverage and availability than “newer” indicators.

- data availability has been improving - for more than 80% of the indicators data coverage improved;
- however, the availability of data still needs considerable improvement - at the end of the 1990s for about one fourth of the indicators the population covered for LDCs and SSA is less than fifty percent;
- finally, “availability” might be exaggerated because for 8 MDG indicators¹ predictive models, based on earlier data rather than current data, are used to generate estimates.

Comparability – data increasingly available but analytically useful?

When data is available is it of high enough quality to be analytically useful? Clearly, differences in definitions and methods of calculation of an indicator can mean reduced opportunities for performance comparisons between regions and for comparisons over time.

Overall 18 of the 48 indicators are affected by comparability issues, and in particular Goals 1,2,3 and 5 (poverty and hunger, primary education, gender equality, maternal health) are especially seriously affected.

Definitional issues

¹ Literacy; infant, child and maternal mortality; measles immunisation; improved water and sanitation; and malaria prevalence.

There are three main definitional problems:

1. Lack of clarity in the definition provided by the international agencies. Newer indicators tend to be subject to more debate and revision than the old and established indicators. Frequent revisions results in poor comparability. Examples include “HIV prevalence among 15-24 year old pregnant women”; the “condom use rate of the contraceptive prevalence rate”, and more generally, indicators that measure the spread of HIV/AIDS,
2. Clear international definition, but some countries and some agencies still use alternative definitions to measure the same indicators. For example, in the case of “literacy”, UNESCO provides a very clear definition that is not adhered to in many censuses and indeed in widely used Demographic and Health Surveys (DHS).
3. Clear definitions but difficulties in operationalising them in some country-specific contexts. This problem applies particularly to five indicators² where precise international classification and differences in the strictness of interpretation promote comparability problems. For example, in the case of “the proportion of births attended by skilled health personnel” although international definitions clearly define skilled health attendants as doctors, midwives and nurses, professional categories do not always fall into this precise international categorization.

Methodological differences

For a number of indicators different calculation methods are used to derive estimates. These different methodologies can sometimes introduce biases that severely restrict cross country or time series comparisons. That said, when the methodology is used consistently comparisons can become possible. Some examples are presented below:

- For the indicator “share of women in wage employment in the non-agricultural sector”, available estimates make use of two main sources of information: labour force surveys and establishment surveys (other sources are administrative and official statistics or insurance records). However, the coverage of these surveys makes international comparison difficult. For example, in Algeria the establishment survey covers only the public sector, whilst in other cases establishment surveys can exclude the informal sector or small enterprises,
- For the indicator “share of poorest quintile in national consumption”, about 40% of reported values are income shares instead of consumption shares. This generally tends to produce values that are lower than those calculated with consumption. A similar problem applies to the poverty indicators,
- In the case of measles immunization, when available, household survey data are often used to recalibrate officially reported estimates, but if household survey estimates do not exist, only official data are taken into consideration. For instance in Tunisia, official reported coverage since 1997 was above 90%, but a recent household survey estimated a coverage in 1999 of just above 70%,
- In indicators of access to water and sanitation improved systems, administrative data report a provider perspective while household surveys are closer to the user perspective. Although much of the provider data has been replaced with data from household surveys (whenever possible), there are still cases in which administrative data are the only available source³.

³ If administrative data are the only available source of data there is certainly a benefit in using them, but when different sources of data are used they should be properly distinguished.

Priority activities

Unfortunately, many of these problems can only be addressed as more data become available. However, there are areas that are within the control of international agencies that if properly addressed could lead to substantial improvements in international data even in a relatively short time period. These areas are not based on the collection of more data, but on better management and use of the present resources. The key areas include:

- appropriate and additional use of available data from household surveys with an investigation of the potential benefits of an international household survey database;
- changes in the use by international organisations of data reporting questionnaires sent out to national governments;
- changes in the use of international population data in the calculation of some indicators;
- changes in the management of common methodologies and definitions; and
- changes in data management practices.

Appropriate and additional use of household surveys and the creation of an international surveys database

The use of data from household surveys could dramatically improve both the quality and coverage of certain indicators that do not currently make full use of household survey data. This would potentially apply to four indicators in education and literacy: the net enrolment ratio (NER), the survival ratio, literacy rate of 15-24 year-olds, the ratio of girls to boys in primary, secondary and tertiary education, and the ratio of literate females to males of 15-24 year-olds.

The inclusion of data from existing household surveys could make significant differences to the estimates for these four indicators. For the NER we could have information on an

extra 10% of countries; for the survival rate to grade 5, this percentage increases to about 40%, and in the case of literacy indicators estimates could be based on information that on average is more recent by eight years.

In addition to these indicators household survey data could also potentially be used for 'the proportion of population below minimum level of dietary energy consumption'. For this indicator, household survey data could complement the information currently available and challenge some of the present results with benefits for the quality of the data.

An international surveys database would go some way to provide a sustainable solution to problems of availability, comparability and timeliness. Some of the MDG agencies have put in place sophisticated networks in order to identify and select household surveys for use in generating MDG indicators. Examples include WHO's Global Database on Child Growth and Malnutrition; UNICEF's CRING database that includes data on Infant and Under-5 Mortality Rates. All agencies maintain some type of formal or informal network for that aim to capturing new available sources of data, including household survey data.

However, these networks can and do miss important data gathering opportunities (e.g. malnutrition data in Malawi; mortality and immunisation data in Pakistan) from multi-topic household surveys that collect a range of MDG data. Such multi-topic household surveys are becoming increasingly available for example the MICS, the DHS, and the LSMS.

An international household survey database could potentially document and archive the surveys and collect and eventually provide support documents (questionnaires, manuals of interviewers, sampling information, etc.). The benefits of such a database would be of three kinds:

1. An improvement in data availability and timeliness. Given that almost half of the MDG indicators use or may potentially

make good use of household surveys, creating a system that systematically gathers all household surveys with relatively easy access to data and information could reduce substantially the chances of failing to include available estimates.

In the case of Pakistan, and inclusion of the Pakistan Integrated Household Survey in such as database would have changed infant mortality rate estimates from 84 to 77 for 2001 and from 96 to 108 for 1990.

2. To enable the improvement of comparability across countries. Gathering different questionnaires in one database could help highlight definitional differences and the importance of country specific methodologies. A single database would improve the chances for a harmonization of some definitions.

For example, because of substantial differences in definition, some surveys cannot currently be used. This is the case for categories of water sources and sanitation, and for skilled birth attendants, among others.

3. Reduction in the costs both of reporting MDG performance and of using data for analysis. National statistical agencies currently have to report to several different data agencies. A single international database would require only a single report.

Similarly, distribution and analysis of data from a single source would reduce supplier and user costs.

Changes in the use by international organisations of data reporting questionnaires sent out to national governments

For 12 indicators, international agencies use questionnaires that are sent out to national governments, to gather information from reporting countries. These indicators include

amongst other education, immunisation and improved water and sanitation.

Although questionnaires represent a convenient way to gather information, this approach does not always provide the best means of acquiring high quality data, especially when the questionnaire is long and complicated. Indeed, the authors' own experience and observations are that these questionnaires do not receive the attention that they deserve. Reporting governments do not necessarily prioritise and assure quality control in their completion of the questionnaires.

Alternative ways that could be explored to avoid such problems are a direct contact with the authorities that fill the questionnaires and the independent gathering of similar information using other sources.

Changes in the use of international population data in the calculation of some indicators

International population data are used in two quite different ways in the calculation of MDG indicators:

1. They are combined with national data to produce country indicators;
2. They are used as weights to generate regional or global estimates.

It is when international population data are directly used to produce country estimates that agencies could be more cautious vis a vis possible unwanted effects on some MDG indicators.

In fact, population data for many countries are just estimates and consequently these estimates are surrounded by some uncertainty. There are three main international sources of population data: the United Nations Population Division, the World Bank and the U.S. Census Bureau. Furthermore, most national governments make population estimates and projections for their own countries. These various population estimates present differences that

cannot be ignored - especially for developing countries.

Comparing population estimates of the UN Population Division, the World Bank, and the US Census Bureau we found that for the year 2000, total population estimates presented differences greater than 10% (more or less than 10%) for more than 1 country in every 6, and this percentage increased to almost 1 in every 3 when taking into consideration estimates of population aged between 0 and 14.

In order to mitigate the potential unwanted impacts, where faced with discrepancies international agencies could consider several hypotheses, and investigate the reasons behind the discrepancies. International agencies could also be encouraged to make public several alternative figures at each end of the potential scale of difference, making their sources clear. Additionally, other sources of information such as household survey data could be used to cross-check data where anomalies occur.

Changes in the management of common methodologies and definitions

Indicator definitions are not always accepted or as widely known and understood as might be desirable. This is particularly problematic when there is a lack of data: when data does become available, definitions may have become distorted to such an extent that an accurate trend analysis is impossible.

For example, in the case of "The proportion of births attended by skilled health personnel", skilled health attendants are defined as doctors, midwives and nurses. However, in some countries there are professional categories that do not always fall into this precise international categorization. This creates uncertainty concerning the way in which data is treated. Indeed, this is the main reason for some of the differences between WHO and UNICEF data, especially in Latin American countries. For instance, both WHO and UNICEF rely on the 1998 ENSMI survey for the estimate of births attended by skilled health personnel in Paraguay. While WHO reports a proportion of 58.1, UNICEF's

estimate is 70.9. This apparently is due to the exclusion of 'partera' from the WHO estimate.

Possible responses are: a) to further promote standard definitions and guidelines (this applies to situations where there are distortions in definitions, or a need for guidance in using the definitions, for example in education and HIV indicators); b) to provide tools with which to successfully measure trends in particular countries where context-specific issues arise.

Changes in data management practices

When data is posted on international agencies' websites, original notes and sources can become lost. This means that to the observer, information looks as if it is directly comparable, while in fact it may not be. The use of different methodologies does not undermine the analysis of data trends for countries that systematically use the same source. However, it does raise problems for comparability with other countries that may use different sources.

It is therefore important to ensure that metadata is always published with a link to its source. Certain exceptional data should also be properly documented. For example, in the indicator for 'the share of women in wage employment in the non-agricultural sector', some of the estimates refer only to urban areas of the country. Transparency also dictates that where figures are derived from models, the source and year on which the model was based, should also be clearly presented.

Further Issues

Finally, this study identified a number of areas in which the international effort could focus to improve the quality of data. In particular, the issues that might be addressed in further dialogue with the agencies are:

- a. The level of accountability that the various lead international agencies have in the reporting process, as well as a review of the rules and systems that

define the responsibilities of countries versus international agencies;

- b. A further investigation into possible improvements in the quality of data currently collected from agency questionnaires;
- c. Further exploration of the feasibility, costs and benefits of implementing an official international household survey database;
- d. Explore the advantages of a direct re-analysis of raw survey data.

Contact	Ludovico Carraro
Telephone	+ 44 (0)1865 207300
Email	ludovico.carraro@opml.co.uk