

Paper presented at the Conference on Data Quality for International Organizations, 27-28 May 2004, Wiesbaden, Germany, organised by the Committee for the Coordination of Statistical Activities

DATA INCONSISTENCY, STATISTICAL CREDIBILITY AND THE HUMAN DEVELOPMENT REPORT

Haishan Fu, Chief of Statistics, Human Development Report Office, UNDP¹

Summary

This paper addresses the issues of data consistency and coherence of international statistics. It provides a “user’s eye view” of the different types of data quality issues often encountered by international data users and disseminators, and their implications for statistical credibility and the impact on policy discussions. It discusses the key factors that often lead to data inconsistencies, and describes the strategy that the Human Development Report Offices has developed to deal with these issues. While emphasizing the significant progress the international statistical system has achieved in its continuing efforts to set common standards and establish quality control frameworks, the paper stresses the need to further improve the practices of international data producers, especially in the areas of coordination (both with national statistical organizations and among themselves), promotion of international standards and definitions, and meta data dissemination that enhances the understanding and interpretability of international statistics.

Introduction

“I believe that having access to official statistics which we can all trust is essential in any healthy society... For official statistics to play[that] key role effectively in democracy we need to have confidence in the figures themselves. “

- Tony Blair (1999)

Over the past 14 years, through the publication of the indicator tables in the *Human Development Report* (HDR), the Human Development Report Office (HDRO) of UNDP has become one of the major users and disseminators of international statistics. The indicator tables in the HDR, led by the human development index (HDI), are widely regarded as a useful source of data on human development – they are systematic, comprehensive and easily accessible for public policy discussions and academic research. Because the tables offer cross-country comparisons, they are drawn from the leading international data series that are the best available when the Report is produced each year.

These international statistics also form a statistical basis for the substantive policy discussions in the HDR, which usually focuses on a different, often politically current and conceptually path-breaking, theme of human development each year. Because of its wide reach and clear impact upon national and international development thinking and policy debate, the Report must be based on solid statistical credibility, determined largely by the quality,

¹ The author thanks Sakiko Fukuda-Parr, Tom Griffin, Claes Johansson, Chris Kuonqui and Ricardo Fuentes for their assistance and comments during the preparation of this paper.

consistency and comparability of international statistics it presents, as well as the proper use of these data according to the highest professional standards.

The Human Development Report Office has made continuing efforts over the years to improve its statistical practices, and to achieve a balance between “[the] desire to disseminate as much relevant information as possible in the interests of improving transparency with an equally important concern for development advocacy,” (Ward 2004). With the support and cooperation from many international data agencies and experts, the quality of the Report’s statistical content and presentation has been significantly improved. Yet, several data quality issues, in addition to the existing data gaps in many areas, constantly pose questions on the accuracy and reliability of the international data presented in the Report, and cast doubt on its overall statistical credibility among users of the Report.

Data inconsistency and incoherence – a user’s experience

Three types of data inconsistency are of main concerns to the HDR: inconsistency between national and international data; inconsistency/incoherence between international data series; and the differences in timing and frequency of data revisions among international agencies.

Inconsistency between national and international sources

Inconsistencies between national and international estimates for some indicators, though inevitable in certain cases due to necessary international harmonization, often raise serious concerns among users of the HDR – a large proportion of them being from national governments. Such discrepancies are often brought to light through government complaints to the HDRO: sometimes the data gap in an international data series is contested and a national estimate is claimed to be available; but more frequently, the accuracy of the international estimate is questioned and a different national estimate is proposed.

This data inconsistency problem may undermine the statistical credibility of any global assessments of human development across countries, affect the effectiveness of policy discussions, and weaken the policy dialogues with national governments. But due to its high profile and the political sensitivity and impact of the HDI, the HDR has been subject to serious scrutiny by governments and other users.

While most of the data inconsistency issues have been raised in the context of the HDI, increasing attention is paid to other basic indicators, ranging from education to health to economic statistics. Two contributing factors: (1) the HDR has been more frequently used as an alternative source of statistics for country level analyses and monitoring; and (2) the new level of global and national commitment to monitoring development goals and targets demands greater consistency between national and international data.

Despite much progress in recent years, the data inconsistency problems remain substantial, and they call into question the quality of the international statistics.

Inconsistency/Incoherence in international data series

There also exist inconsistencies and incoherence among international data series in some areas – either between data series produced by different agencies for the same indicator, or among the various data sets based on which other indicators are derived. Here are some examples:

- Due to differences in data sources, timing of updates, and definitions, the population estimates made by the World Bank differ from estimates by the United Nations Population Division (UNPOP) for a significant number of countries (with a difference between 5-30% for over 30 countries and, in the case of Bhutan, the UNPOP estimate is 1.5 times higher than the those of the World Bank estimate). These population estimates may also differ from national estimates and projections. As many international agencies rely on the UN official population estimates and projections for deriving indicators, while others rely on World Bank, these differences are often already reflected in available indicators in other areas.

This leaves the HDRO, as well as many other users, in the position of having to use different population estimates, or indicators derived from them, for the same country in the same report, or even in the same table. One compelling example is the HDI – its components are based on different population estimates: life expectancy, adult literacy and gross enrolment use UNPOP estimates, while GDP per capita uses World Bank. While technically we could adjust the data to reflect consistent population estimates, discrepancies between the resulting estimates shown in the HDR and the original published data elsewhere would create further confusion. The problem is not limited to indicators, but also aggregates based on population weights.

- Some of the very basic education indicators, such as enrolment ratios and other indicators derived from them (e.g., school life expectancy), are calculated on the basis of national administrative records (numerator) and the official UN population estimates (denominator). Depending on the potential differences between the actual size of the population and the international estimates, the resulting enrolment ratios could be distorted, and in the extreme case, the estimated net enrolment ratios are shown above one – the theoretical upper limit – for a number of countries.
- The World Health Organization has been advocating the usefulness of a health adjusted life expectancy (HALE) measure. The HDR has featured this innovative approach recently but could not proceed to include the data in the Report. The reason? The HALE estimates are derived on the basis of a set of life tables developed by the WHO independently, which are not consistent with the UNPOP produced official life expectancy estimates presented in the HDR. For some countries, the estimated HALE is actually higher than the life expectancy estimates from UNPOP.
- Though efforts have been made by UNSD and the International Energy Agency to improve the consistency between their data series on energy consumption (e.g., electricity), large differences still remain for a number of countries.

Similar issues exist in other areas of statistics. Though less visible perhaps than the issue of inconsistency between national and international data, these inconsistencies and incoherencies in international data series reflect the quality of underlying data, undermine statistical credibility of international statistics, and deserve serious attention.

Different timing and frequency of data revision

Revisions to data usually represent further improvement of existing series based on more recent information available or refined methodology. But the differences in the timing and frequency of data revision by various agencies can also pose distinct challenges to reports like the HDR.

For many international data series, data revisions are carried out regularly (e.g., biannual revision of population data by UNPOP and yearly updates of economic data by the World Bank). In such cases, the challenge for the HDR is to explain the revisions to its users, who often interpret any drastic year-to-year changes in estimates published in consecutive editions of the Report as errors, or question why the “history is rewritten” and doubt the credibility of the data. With the continuing cooperation of many agencies and the increasing availability of metadata, our job has been made much easier, though it remains difficult.

Revisions that are frequent and large may easily disrupt databases and cause inconsistencies. Some agencies, striving to constantly improve the quality of their data, have adopted a “continuing updating and releasing” policy – as soon as new information becomes available from a country, a revised estimate is introduced in the existing data series. This often means that, as soon as the HDRO published its Report, revised estimates from the source agency are issued. The inconsistency between the HDR and the source agency data is another source of complaints by governments and other users.

Differences in timing of revisions among agencies that disseminate data on similar indicators or data that feed into other agencies’ data series can introduce additional incoherence in international statistics. For instance, the fact that other estimates and projections such as economically active population and urban population often lag behind updates in population estimates, makes it inevitable that these estimates are not consistent with the population estimates shown in the report.

Why do inconsistencies exist? And what is required to reduce them?

As Giovannini and Ward (2004) point out, “the quality of statistics disseminated by an international agency depends on two aspects: the quality of national statistics received, and the quality of internal processes for the collection, processing, analysis and dissemination of data and meta data.” The following factors appear to be key to the existence of data inconsistencies and require strategic priority.

Quality of national statistics. Some of the differences between national and international data are inevitable – they can be a result of the very process of international harmonization, in which national data, inconsistent with the international standards and definitions or of poor quality for other reasons, need to be adjusted. Where data for a country are missing, international agencies may also produce an estimate if other information can be used. In some cases, the international indicator (such as GDP per capita in PPP US\$) is produced mainly for international comparisons and is not directly comparable to other related national statistics.

Building national statistical capacity remains a pressing challenge for the entire international community. While national statistical development strategy should be based on national priorities, it is imperative that, to the extent possible, internationally agreed standards and definitions in basic areas of statistics should be promoted and adopted by countries.

The international statistical community is not only challenged to continue promoting the adoption of existing standards and definitions, but to play an active role in developing new approaches to meet urgent needs (e.g., comparable measures of literacy).

Coordination between national and international agencies. Other data discrepancies between national and international sources exist because of the difficulties in the coordination between

national statistical systems and international agencies: the most recent national data may not be made available in time for international agencies to incorporate in its updated data series.

Improving coordination between national and international data producers (including regional statistical agencies) remains a constant challenge to international statistical community. Most international agencies recognize the significant implications of data inconsistency and seek to reduce the problem through improved work programmes (such as regional workshops by UNESCO Institute for Statistics) to improve data collection from countries, and through country consultations (e.g., WHO and UNESCO Institute for Statistics) to ensure data consistency.

Related to the national-international coordination is the issue of *coordination within country* among different statistical agencies, which is essential to achieve consistency and efficiency in the international statistical system. The HDRO is frequently contacted by national government agencies who seemingly have no clear idea how the international data series are compiled and what the original national sources are.

The international statistical community should not shy away from these questions: To what extent could the national statistical office become a central coordinator within a country in coordinating with international data agencies? And how can the international community help develop the technology to facilitate the establishment of a central data repository in developing countries that enhances the dissemination and use of existing data?

Coordination among international agencies. Improving the consistency and coherence of international statistics requires improved coordination among international agencies, which, in turn, requires will, commitment and innovative approaches from all involved. Success is possible, as evident in the areas of trade statistics (UNSD 2003) and data on infant and child mortality produced by UNICEF/WHO and the World Bank. Other new initiatives also offer much promise, such as the Health Metrics Network (WHO 2004) and the Literacy Advisory Board set up by the UNESCO Institute for Statistics.

But much more is needed, both in the processes of data collection and compilation/dissemination, if we are to reduce the burden on countries, to obtain better and more timely data from countries, and to achieve more coherent and consistent international statistics in various areas. This Committee has a particular responsibility to promote the formation of specific coordination mechanisms among agencies producing data on the same indicators, or between agencies whose data feed into the data series of the other agency.

Uncertainty of estimates. Inconsistency or incoherence in international statistics is often a reflection of the high level of uncertainty of the underlying data. One background study for the HDR 2004 (Scott 2004) revealed that, in the case of Kazakhstan, depending on the source of information used, estimates of infant mortality published by four different regional and international agencies could vary from 19 to 81 per 1,000 live births. The level of uncertainty in estimates can have profound impact on policy assessments, especially trend analyses.

Guided by the *Fundamental Principles of Official Statistics*, and in the interest of data users (and their understanding of the international statistics and the possible inconsistencies), the international statistical community should consider promoting the publication (perhaps along with meta data) of margins of uncertainty for key statistical indicators. Particularly concerning trend analysis, such indications from statistical agencies of how accurate preliminary or estimated data are would be useful to researchers, as they may decide not to use these data in their time series if they are outside a certain degree of accuracy (Carson 2003).

Meta data dissemination. Much of the concern of international data users over the data consistency issues is resulting from a lack of understanding of the processes and methodologies related to the production of specific international data series. To enhance users' understanding, and to ensure transparency, international data agencies need to make sure that meta data are disseminated and made easily accessible. Informed users, especially national governments, could, in turn, directly influence national data production and coordination with international data producers, and thus help improve the quality of international statistics.

The international statistical community has made important efforts to encourage meta data dissemination by international agencies. It is now time to fully develop the international guidelines for the presentation of statistical data and metadata, a task currently underway under the leadership of OECD (Ward 2003).

Revision policy. As various revision practices contribute to the inconsistency and incoherence of international statistics, it is important to form sound revisions policy—an essential aspect of strengthening governance of statistical systems (Carson 2003) and has wider implications—at international level.

Apart from the regular updates and revisions, it is particularly important when international estimates need to be revised as errors in reporting and processing are uncovered. In such cases, the revision should be made in a transparent and timely manner, and “to lessen the trauma caused by revision, users would want clear documentation” (Carson 2003).

Dealing with data inconsistency and statistical credibility – the HDRO strategy

The increasing emphasis on international and national monitoring of progress toward development goals and targets demands better national and international statistics that are consistent. The HDRO has found itself, more frequently than before, in a difficult situation of having to explain the inconsistencies and the year-to-year revisions, and having its credibility challenged. It has worked increasingly between international data agencies and national governments, seeking clarifications, providing technical explanations and facilitating communications and coordination between the two.

Many of these cases have turned out to be constructive for improving data quality and understanding. There are cases where weaknesses in international series were discovered and corrected, and cases where governments (as users of the HDR) were linked directly to the international agencies and its own data producers, and helped improve the data availability/transfer to international agencies. But clearly, a more systematic approach is required.

In a recently developed strategy to deal with data inconsistency and its implications for the HDR credibility (HDRO 2003), the Human Development Report Office looked into key factors relating to both the data production and dissemination processes, and HDRO's unique role/position in each. It proposed two strategic targets: promoting coordination among producers and promoting understanding among users.

Promoting coordination among producers

The HDRO recognizes that, as primarily a user, not a producer, of international statistics, the HDRO's direct role in the data production process is limited. But given the Report's political

influence (and the impact of the HDI), HDRO could and should campaign actively for the transparency and accessibility of international agencies' methodology and procedures, and the development and adoption of international standards and methodologies by countries. The HDRO could also help promote the coordination between national and international data producers, not only through advocacy but also by becoming more actively involved, or being more systematic, in the communication process between the two.

Working through the UNDP country offices. In particular, through UNDP country offices and around the time of the HDR launch (when attention to its data is at its peak), the HDRO could identify the problematic data areas that governments have concerns with, communicate the issues to the respective international data agencies, and help follow through the process to ensure the data problems are clarified or resolved.

The UNDP country offices are not only users of the HDR, but partners in outreach/dissemination of the Report, they are also the institutional link between UNDP/HDRO and the country governments. Through direct policy dialogues and various working programs with the governments, including statistical capacity building activities, the UNDP country offices are in a better position to communicate to/from the governments and encourage them to collaborate more effectively with international data agencies.

This is to be carried out with the objective of supporting and complementing the existing data collection/country consultation processes established by international data agencies, not interfering with them.

Establishing formal procedures with data agencies. HDRO also aims to establish formal coordinating procedures with key international data agencies so as to deal effectively with specific data problems raised.

Over the past few years, the HDRO has formed close collaboration with a number of international data agencies, including the UNESCO Institute for Statistics, UNICEF, UNPOP, WHO and the World Bank. These agencies have provided extensive and timely technical support to HDRO, especially in cases when data consistency is being questioned; HDRO has also assisted some of the agencies to gain access to the more recent data from countries. We wish to formalize this collaboration by setting up a set of procedures to ensure that agencies are informed of any relevant data issues in time, committed to follow up with proper technical advice/clarifications; and HDRO will provide further assistance through UNDP country offices in dealing with country governments/national data producers.

Promoting understanding among users

Feedback from users of the HDR often indicates that many of them, including national governments or even national experts, have surprisingly little knowledge about the international data production process. Many people have long-held misconceptions about the statistical principles and contents of a global report. Some governments do not seem to be fully informed of the regular procedures for incorporating national data into international series. When data gaps or inconsistencies are uncovered, governments often demand that HDRO collected data directly from countries or send data to HDRO directly for inclusion in the Report.

The HDRO is seeking ways to enhance the understanding of users of the Report through an educational campaign. Some of the planned activities include improving data dissemination and information outreach methods, developing a new HDRO statistical website, conducting

workshops with country offices, and introducing thematic consultation in the regular process of HDR consultation with member states through the UNDP/UNFPA Executive Board.

Conclusion

This paper offers a user's view on the issue of inconsistency and incoherence of international statistics, and its impact on the statistical credibility of global reports that aim at monitoring progress in human development. It identifies some of the key factors that contribute to the data quality problem and suggests several concrete approaches to improve the current situation. It also describes the strategy the Human Development Report Office has adopted to deal with the issue in disseminating international statistics.

It recognizes that, over the past decade, the international statistical system has evolved quickly in response to the increasing demand for better statistics to manage and monitor human development. Much effort has been spent, and good progress has been made, in setting new international standards and quality control frameworks to improve the availability and quality of international statistics.

But much remains to be done in establishing frameworks and improving guidelines. Though the process is complex and changes often require a more fundamental shift, a lot can still be achieved based on common principles and commitment.

In particular, this may wish to consider promoting even stronger But much remains to be done in an established framework and improved guidelines. Though the process is complex and changes often require more fundamental shift, a lot can still be achieved based on common principles and commitment.

In particular, this Committee could adopt as a priority objective to further promote better coordination between international data agencies and country statistical authorities, and encouraging the formation of specific coordination mechanisms among international agencies that disseminate data on the same subject, or when data series from one agency forms the basis of data for another. (The Human Development Report Office is willing to participate in documenting the best practices in this area.) It may step up its efforts to develop, and put into effect, clear guidelines on meta data dissemination, including the publication of uncertainty range of estimates. It should also actively pursue the establishment of revision policy and guidelines for international data.

END

References

Blair, Tony. 1999. Introduction to "*Building trust in statistics—The White Paper on Statistics.*" [<http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=10543&Pos=&ColRank=1&Rank=272>]. Accessed May 2004.

Carson, Carol S. 2003. "*Revision policy for official statistics: A matter of governance.*" Committee for the Coordination of Statistical Activities, Second session. Geneva, 8-10 September 2003. SA/2003/13. [<http://unstats.un.org/unsd/acsub/2003docs-2nd/sa-2003-13.pdf>]. Accessed April 2004.

Giovannini, Enrico and Ward, Denis. 2004. "*Quality framework for OECD Statistics: Getting our own house in order.*" Committee for the Coordination of Statistical Activities, Conference on Data Quality for International Organizations. Wiesbaden, Germany, 27-28 May 2004.

HDRO (Human Development Report Office, UNDP). 2003. "*Addressing the discrepancies between national and international data: A strategy to improve the statistical credibility of the HDR.*" November. New York: HDRO.

Scott, Wolf. 2004. "*Uncertainty of estimates.*" Background note prepared for the *Human Development Report 2004*. Geneva.

United Nations Statistics Commission. "Fundamental Principles of Official Statistics." [<http://unstats.un.org/unsd/methods/statorg/fp-english.htm>]. Accessed April 2004.

UNSD (United Nations Statistics Division). 2003. "*Coordination of data sets disseminated by different organizations covering the same subject matter.*" Committee for the Coordination of Statistical Activities, Second session. Geneva, 8-10 September 2003. SA/2003/18. [<http://unstats.un.org/unsd/acsub/2003docs-2nd/sa-2003-18.pdf>]. Accessed April 2004.

Ward, Denis. 2003. "*Proposal for the development of international guidelines for the presentation of statistical data and metadata.*" Committee for the Coordination of Statistical Activities, Second session. Geneva, 8-10 September 2003. SA/2003/11. [<http://unstats.un.org/unsd/acsub/2003docs-2nd/sa-2003-11.pdf>]. Accessed April 2004.

Ward, Michael. 2004. *Quantifying the World – UN Ideas and Statistics*. United Nations Intellectual History Project Series. Bloomington, Indiana: Indiana University Press.

WHO (World Health Organisation). 2004. "*The Health Metrics Network: An emerging global partnership for health information.*" Background note prepared for the *Human Development Report 2004*. Geneva.