MILLENNIUM DEVELOPMENT GOALS SYSTEMATIC TIME-BOUND EVALUATION OF PERFORMANCE AND SYSTEMS (STEPS) INDICATORS

October 2009

To provide a more complete picture of progress toward the Millennium Development Goals (MDGs) and allow countries to better focus their efforts, it is proposed that indicators of national performance at achieving the MDGs be developed. These indicators would supplement the existing data on MDG targets and the various lists of output indicators by focusing on the process and inputs for achieving the goals.

Country performance, or country effort as it is also termed here, is often assessed with regard to results, whether outputs or eventual outcomes. But if performance falls short, it has to be examined directly, not just through its consequences. If performance is to be improved, one has to know more about it, rather than just measuring the results.

For instance, to improve education outcomes (in relation to MDG 2), it helps to have data on such educational outputs as net primary enrolment and the proportion of students who complete primary school. But it is similarly important to have actual measures of performance relating to such matters as (a) appropriate governmental emphasis on primary education, (b) sufficient and timely funding, (c) the existence or construction of the necessary school buildings, (d) the provision of appropriate textbooks, (e) the removal of barriers to enrolling girls, etc.

Indicators for such aspects of performance have been produced in a timely, cross-national fashion for some years in limited areas relating to the health sector. It is proposed that the approach, which relies mainly on expert ratings, be extended to other MDG sectors, social, economic, and environmental. Performance indicators--to be referred to here as STEPS indicators, for Systematic Time-bound Evaluation of Performance and Systems indicators--should allow better understanding of why particular countries are moving toward the MDGs more rapidly while others are falling short. Identifying the shortcomings of particular programs and the imbalances within them through STEPS indicators could help accelerate progress toward the MDGs.

CONTENTS

ACRONYMS AND ABBREVIATIONS

AIDS	acquired immune deficiency syndrome
API	AIDS Program Effort Index
CAP	country action plan
CAS	country assistance strategy
CCA/UNDAF	Common Country Assessment /UN Development Action Framework
CIESIN	Center for International Earth Science Information Network
CO2	carbon dioxide
DAC	Development Assistance Committee
ESCAP	Economic and Social Commission for Asia and the Pacific
GDP	gross domestic product
HIPC	heavily indebted poor country
HIV	human immunodeficiency virus
MDG	Millennium Development Goals
MDRI	Multilateral Debt Relief Initiative
MNPI	Maternal and Neonatal Program Index
ODA	official development assistance
OECD	Organisation for Economic Co-operation and Development
ONE-DATA	One-Debt AIDS Trade Africa
PPP	purchasing power parity
STEPS	Systematic Time-bound Evaluation of Performance and Systems
UN	United Nations
UNAIDS	Joint United Nations Programme on HIV/AIDS
UNCT	United Nations country team
UNDESA	United Nations Department of Economic and Social Affairs
UNDP	United Nations Development Programme
UNIFEM	United Nations Development Fund for Women
USAID	United States Agency for International Development
WHO	World Health Organization

INTRODUCTION

Millennium Development Goal (MDG) 8 calls for a global partnership for development in pursuit of the MDGs. This partnership must be grounded, at the country level, in government efforts to reduce poverty and develop human resources, allied with efforts of private organizations and individuals. It requires financial, technical, and policy support from bilateral donors, multilateral institutions, and new sources of development finance such as foundations. To be effective, the partnership needs effective multilateral institutions, rules-based policy frameworks, and the cultivation of regional, bilateral, and private-sector relationships.

All these elements must be monitored, and many of them are. Outcomes are tracked. The annual report on MDG progress (United Nations 2009b) provides a comprehensive stocktaking across MDGs 1 through 7. For MDG 8, donor inputs are tracked. The report of the MDG Gap Task Force (United Nations 2009a), which has become an annual publication, assesses donor progress. Other indices and reports, such as the Commitment to Development Index (Center for Global Development 2009) and the annual ONE-DATA report (ONE 2009) on the fulfillment of commitments to Africa, also provide broad assessments of donor performance. Taking all these reports together, however, still leaves a lacuna in tracking the efforts of national governments.

This paper proposes to address this "missing middle" by developing indicators of national performance in relation to each MDG. In line with the 2005 Paris Declaration's commitment to mutual accountability in the use of development assistance (OECD 2005), indicators of performance are meant to help focus UNDP, the United Nations system, and other donor assistance on building in-country capacities where weaknesses are identified. The approach is also intended to determine what developing countries have invested in inputs and processes and to highlight these efforts, so that, if outcomes still do not match, investigation is possible and the case can be made for the right kinds of further assistance.

The performance indicators proposed here will be referred to as **Systematic Time-bound Evaluation of Performance and Systems (STEPS) indicators.** They are designed to allow for straightforward and frequently updated assessments of country performance, in a cross-nationally consistent fashion, relating to MDG goals and targets. The simplicity of obtaining the indicators is intended to provide data of use for the 2010 review of MDG progress and inform adjustment of country strategies and approaches where necessary.

To provide these indicators, once collected, to the international community as a whole, it is proposed to use an **online dashboard**, a device to make results readily and easily available to everyone—governments and national agencies, United Nations agencies, the donor community, civil society organizations—in a form in which results can be easily visualized and the relative importance of different indicators can be studied.

An illustration of the STEPS approach is provided in this proposal and its annexes, focusing on one of the MDGs. A standard questionnaire and a potential schedule for refining it, collecting data, and analyzing them are provided. It is intended that national governments work with key stakeholders and partners in interpreting the data and drawing appropriate implications about how to accelerate progress toward the MDGs.

The country MDG STEPS indicators will complement existing MDG monitoring conducted by UNDESA and the MDG monitoring tools by:

- Providing a snapshot of a country's overall performance and the effort being exerted relating to specific MDGs

- Helping identify bottlenecks that require further attention

- Indicating in advance where progress toward the goals is most likely because of efforts already being made

- Assisting United Nations Country Teams (UNCTs) in identifying areas of insufficient capacity, where additional support is needed

- Providing a basis for achieving consensus among stakeholders and line ministries relating to specific policy interventions

- Providing data to use in advocating particular policy changes

- Simplifying and standardizing presentation of data relating to performance and outputs, providing simple visualizations and encouraging utilization

- Enhancing the ability of decision makers, development practitioners, and analysts to make cross-country comparisons and draw useful inferences

- Within UNDP, complementing the approach to assessing national efforts (as defined by the Common Country Assessment /UN Development Action Framework [CCA/UNDAF] guidelines) and enhancing their alignment with the MDGs.

This proposal lays out the case for performance or effort indicators to contribute to all of these goals. First, a brief situation analysis is presented, illustrating the need, already evident to many, for accelerating progress toward the MDGs. The situation analysis also discusses the MDG indicators currently available and reviews experience with performance indicators in particular fields. Second, the proposal lays out a strategy for developing performance indicators, from design and data collection to statistical analysis and dissemination. Details of the approach are provided in several annexes. Third, the proposal lays out a timeline for the work and indicates the needed resources.

Achieving the MDGs by 2015 continues to require considerable effort at every level, from international donors to developing country governments to provincial and local governments, as well as from civil society organizations and the private sector. Whether these goals will be met in every country is still hugely uncertain. If more effort and better sector performance are required, it would be of great utility to be able to track performance and effort and to assess what difference they are making.

This section illustrates the progress, or lack of it, being made toward the goals, notes the steps needed to move faster toward them, outlines how progress is now being monitored, and considers the case for further indicators of the process that focus on performance.

Insufficient progress

The Millennium Development Goals Report 2009 notes how the economic crisis has had grim repercussions, probably stalling progress toward the MDGs. Even before the crisis, progress was problematic. The 2008 report called for greater effort from governments, donors, civil society and the multilateral system. Though the report proclaimed that the "mid-point shows some key successes" (United Nations 2008:4), it noted areas where more attention was needed. Statistics provided in the 2008 report illustrate why greater effort is needed.

For example, in education and in water and sanitation, the prospects are somewhat ambiguous, with substantial progress but also a threatening shortfall in some areas and regions.

- Net primary enrolment in developing regions, estimated at 80 percent in 1990/1991, rose to 88 percent in 2005/2006. A linear projection of this rate suggests that net primary enrolment could reach 93 percent by 2015, just short of the developed region level of 96 percent in 2005/2006. A little more effort might be enough to reach 96 percent, which could be considered to represent the aggregate goal of "universal primary education." However, sub-Saharan Africa lags behind, with a rate projected to reach only 82 percent by 2015.

- The proportion of the population of developing regions using an improved drinking water source, only 71 percent in 1990, reached 84 percent by 2006. This is just short of the goal of cutting the proportion not served in half (which would imply 86 percent with an improved drinking water source), which should be achievable by or before 2015. However, the situation is more problematic for sub-Saharan Africa and Oceania, each of which has a 2015 goal of about 75 percent but are likely to reach only 63 and 49 percent respectively.

- A goal of 50 percent reduction in those not served has also been set for improved sanitation facilities. Developing regions as a whole are falling short, with the proportion served rising from 41 percent in 1990 to only 53 percent in 2006. At current rates of improvement, Oceania, sub-Saharan Africa, Southern Asia, and Western Asia will all fall short of the 2015 goal.

In nutrition and health, the situation is, if anything, even less promising.

- Children who are underweight were estimated at 33 percent of those under age five in 1990 and 26 percent in 2006. One may assume that the target of halving the proportion of people who suffer from hunger implies halving the proportion of children who are underweight. Then this proportion should fall to 17 percent by 2015. The current pace of reduction is too slow overall, though some regions--particularly Northern Africa, Latin America and the Caribbean, and South-Eastern Asia--are on pace to reach the target.

- Under-five mortality, having fallen in developing regions from 103 deaths per thousand live births in 1990 to 80 per thousand in 2006, is on track to reach 67 per thousand by 2015, well short of the goal of a two-thirds reduction (to 34 per thousand). Some developing regions are doing better than others, but sub-Saharan Africa and Oceania in particular have fallen well off the desired pace.

- Maternal mortality in developing regions was estimated (very roughly, because good data are not available) at 480 deaths per 100,000 live births in 1990, and at 450 deaths by 2005. At this rate of reduction, the target of 120 deaths per 100,000 by 2015 (a three-quarters reduction) looks well beyond reach. While various developing regions have achieved substantial reductions, only Eastern Asia is on pace to reach its target.

For other MDGs, targets are often not precise enough to allow similar comparisons, but the situation is not any better, according to the 2008 report. Achieving "full and productive employment and decent work for all," for instance, is a "distant possibility" (United Nations 2008:8).

One other area is important to note. Official development assistance (ODA), which reached a historic high in 2005, fell 2.5 percent in 2006 and another 0.7 percent in 2007. The 2007 decline was actually equivalent to a decline of 8.4 percent when prices and exchange rates are taken into account (United Nations 2008:44). In relation to gross national income, the 2007 ODA figure was 0.28 percent of gross national income, substantially below an estimated requirement of 0.54 percent by 2015, according to the UN Millennium Project (2005a:252), or the currently requested amount of 0.7 percent by 2015 (United Nations 2003).

Various reports provide much more detail and depth that explains how such statistics as these should be interpreted, and the 2009 report suggests that things may have gotten worse in several areas. It is clear that there is good basis for caution, if not skepticism. It has certainly become more difficult, rather than easier, to reach 2015 goals. With global

economic contraction, both the will and the capability to tackle these problems is in question in both developed and developing regions.

Even were these goals to be reached, of course, much work would remain to be done in the sectors on which the MDGs focus. Some of the goals are in a sense minimal goals. The goal for under-five mortality, for instance, is six times the current rate of under-five mortality in developed regions; the goal for maternal mortality is 13 times the ratio in developed regions.

Accelerating the pace

What needs to be done to accelerate progress toward the MDGs is probably no different from what has been recommended as appropriate strategy in the past. A systematic planning process is necessary: (a) to explore the key dimensions of extreme poverty in a given country; (b) to determine the public investments that would ameliorate the situation and achieve the MDGs; and (c) to develop phased plans for action ultimately translatable into actual budgets (UN Millennium Project 2005a:57-58).

Planners may stop at this point, but what happens afterwards can be even more critical. Funds have to be disbursed and spent appropriately, and the activities they support have to be properly executed. Implementation of projects can involve many levels of government: besides national governments, regional or provincial governments, local governments, neighborhood groups, as well as civil society organizations, the private sector, and the public at large, whose understanding and support are often critical.

National governments have broad responsibility to set the stage for moving toward the MDGs and must also ensure implementation capacity at all other levels. International donors have to ensure an adequate level of aid to supplement domestic resources, aid provided on a predictable schedule within the context of appropriate national plans. Local governments and local institutions are often at the center of implementing these plans. Their capacity to reach individual households and to mobilize individual commitment is essential. The participation of the private sector is vital as well, given its role in raising productivity and incomes, generating jobs, developing infrastructure, and providing expertise in logistics and service delivery. The public must also be mobilized through appropriate communications to demand services and utilize them in the right way.

Both long-run interventions and "quick wins"--time tested strategies that require only limited additional infrastructure--are recommended. The latter include, for example, annual deworming of schoolchildren in affected areas and community support for tree planting to provide "soil nutrients, fuelwood, shade, fodder, watershed protection, windbreak, and timber" (UN Millennium Project 2005a:57-58).

Has this process--careful planning involving all stakeholders, followed by implementation across multiple levels requiring the mobilization of many participants--worked, and will it work to accelerate the pace toward the MDGs? A precise and reliable answer is really not possible. The process, as described by the Millennium Project, is based on the best current research across a wide range of areas and draws on "best practice" lessons from extensive development experience. The array of indicators used in monitoring indicates that results have been produced in a variety of areas in different countries. But whether this can be attributed to the process, and what elements of the process are critical, is difficult to tell because the process itself--as opposed to the results it is meant to produce--has not been subjected to systematic monitoring.

Indicators of progress

Progress toward the eight MDGs appears to be heavily monitored. The eight goals are operationalized in 21 specific targets for developing countries and the international community. To monitor progress toward these 21 targets, a total of 60 "indicators for monitoring progress" have been listed (see Table A1 in Annex A). However, these indicators provide an incomplete picture of what has been done in pursuit of the goals and targets, and what needs to be done to accelerate progress.

The goals represent broadly defined outcomes, such as to "Achieve universal primary education" (MDG 2), and the targets are more precise definitions of these outcomes--in this case to "ensure that, by 2015, children everywhere, boys and girls alike, will be able to complete a full course of primary schooling." Outcomes may be understood as the desirable end states for the population or society. They can be distinguished from outputs, understood as the immediate results produced by a project or intervention. To a large extent, the 60 indicators represent such outputs. In the case of MDG 2, for instance, these output indicators are the primary school net enrolment ratio, the proportion of first graders who eventually complete primary schooling, and the literacy rate at ages 15-24 years.

Outputs and outcomes provide a picture of where a country stands in relation to the MDGs but do not show what is being done to achieve them. A distinction is commonly made in analyzing projects (e.g., ESCAP 2003:5) between inputs, process, outputs, and outcomes, with one leading to the other (Figure 1). Countries are strongly encouraged to focus on goals and targets (outcomes) and advised to pay attention to indicators for monitoring progress (outputs). Statistics on these are maintained and regularly evaluated. But attention to inputs and process is much less systematic.

Inputs constitute the resources required to move toward a goal: financing, staff, infrastructure, etc. Process, in contrast, constitutes the activities to be carried out, often reflected in the intermediate products of these activities. Input and process are linked iteratively, with some inputs being necessary for a process that generates inputs for the next stage: for instance, financing and facilities are necessary for training, which produces the staff to work on a project or intervention. Given this linkage, both process and inputs are covered together here under the term performance or effort.

Some aspects of performance, to be sure, tend to be closely tracked, particularly financing from donor agencies. In addition, UNDP (2009) recommends a thorough assessment, as part of a country analysis, of the policy framework, the financing framework, and service delivery on the ground. Detailed questions are proposed on these topics. For the policy framework, the questions concern whether the MDGs are integrated with national development strategies and whether local plans are properly designed. On financing, questions concern resource mobilization, budgets, and

expenditure plans, as well as coordination with donors. For service delivery, questions are posed regarding personnel and infrastructure, the quality of services, equitability of access, and monitoring and evaluation.

Particularly where service delivery is concerned, the questions are generic, meant to apply across seven of the eight MDGs (the exception being MDG 8, the global partnership for development). To be useful in a particular country analysis of a specific sector, the questions require much further specification: of the kinds of personnel needed,

left because performance indicators are not systematically collected, the key question is whether such indicators can in fact be produced in a timely fashion.

Performance indicators

Performance or effort has been measured previously across developing countries, using a reputational approach, in the areas of family planning and, more recently, maternal and neonatal health and HIV/AIDS. In family planning, initial indicators were produced in 1972, using a questionnaire

Figure 1. Performance as an aspect of movement toward goals
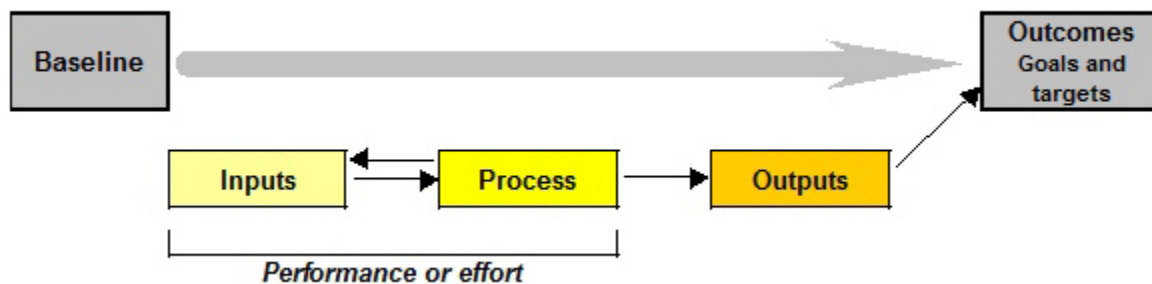
the proper measures of quality, etc. Results are meant to promote corrective action in the specific sector. They are not intended to be compiled across countries or across sectors within one country. With the more specific questions that would be needed, the results would in any case not be easily comparable across countries.

These questions need to be complemented with potentially similar but systematic and uniform performance or effort indicators for use across countries. The lack of comparable effort indicators hampers understanding of each particular situation and makes movement toward the MDGs potentially less easy and transparent. As a recent UNDP-UNIFEM report has noted, "Accurate problem identification requires process-oriented indicators" (Corner and Repucci 2009:11). With the lack of comparable data on effort, public assessments across countries are too often black and white, reflecting success or failure but not the shifting patterns that suggest where countries have begun to go forward and may be close to breakthroughs, or alternatively where they have retrogressed and could soon expect deteriorating conditions. The lack of comparable data means that country assessments are highly individualized and are more easily shaped by potential biases among those participating in the assessments. It means that these assessments cannot learn as easily from the experience of other countries that have moved through similar stages.

At a more basic level, lack of comparable performance indicators impedes global monitoring and evaluation. Global recommendations have been made for how to accelerate progress toward the MDGs. Despite the solid research that generally has gone into them and the accumulated best practice lessons they represent, it is worth having actual evidence that the process is working. Performance indicators would help exclude alternative hypotheses--that perhaps conditions improved for extraneous reasons, or deteriorated because of global conditions rather than any failure in project design or service delivery. Continuing to monitor performance is therefore essential and could turn up further lessons about how to accelerate progress. Given the gaps

developed by leading analysts of family planning programs, R.J. Lapham and W.P. Mauldin (1972). Beginning with the second administration of the questionnaire in 1982, effort data were collected roughly every five years, and the seventh round of data collection is currently in progress.

These data have been collected using a consistent procedure. The questionnaire (mostly unchanged over time) is sent to a small set of expert observers in each country, who rate family planning program effort on 120 items. Combining scores leads to 30 indicators of family planning effort. Raters are carefully selected by the survey supervisors, based on extended consultations, when necessary in country, with those familiar with the country program. Though raters make personal ratings, they are encouraged to consult available data on the country program where these may be useful.

The scores produced by this process have become widely accepted within the family planning community as representing reasonable and comparable measures of the effort that particular programs are putting into achieving their goals. The scores have also proved useful for a variety of purposes. They have provided guidance in allocating support within programs and also across countries, where donors are concerned. They have been useful in making projections, for planning purposes, of the course of contraceptive use or fertility over time (Berelson 1978; Mauldin and Ross 1994; Ross and Stover 2000). They have been critical in allowing analysis of the effects on fertility of family planning programs. That fertility has been successfully reduced in many developing countries over recent decades would have been clearly evident even without family planning program effort scores. But without these scores, it would have been difficult to systematically address the question of whether family planning programs in fact were contributing to fertility or whether socioeconomic development by itself was producing decline. Both sides in this debate have relied to some extent on effort scores (e.g., Pritchett 1994; Bongaarts 1994). A list of 11 applications of family planning effort scores in investigating different topics, and appropriate references,

are provided by Stover (1999). Scores have been sufficiently useful so that subnational scores have occasionally been collected, as in Vietnam (San et al. 1999).

Performance scores for maternal mortality and neonatal health programs are a more recent development. The first set of scores across developing countries (on an instrument labeled the Maternal and Neonatal Program Index, or MNPI) was collected in 1999, following a methodology basically similar to that for family planning. An expert consultation was convened in developing a questionnaire, and raters were selected country by country, based on their expertise, to answer the questionnaire (Bulatao and Ross 2002). Two further rounds of the survey have been conducted, in 2002 and 2005.

These effort scores have been useful for advocacy: "to raise awareness of maternal health issues and help policymakers and program managers identify strategies for reducing maternal morbidity and mortality" (Policy Project 2009). They have also been used in the analysis of the effect of programs (e.g., Bulatao and Ross 2003), though analysis has been less extensive than for family planning effort for a number of reasons. In the area of fertility and family planning, the quantitative orientation is much stronger. For maternal mortality, outcome measures are often unreliable and do not support extensive analysis.

A similar instrument relating to HIV/AIDS programs (the AIDS Program Effort Index, or API), was applied cross-nationally in 2000, 2001, and 2003 (USAID, UNAIDS, WHO, and the Policy Project 2003; Stover et al. 2000). As the 2004 Report on the Global AIDS Epidemic (UNAIDS 2004) reported, program performance improved cross-nationally on each of four major dimensions over these years. Policy and planning scores were generally the most satisfactory, followed by political support, care and treatment, and finally program resources. Subsequently, UNAIDS has placed more emphasis on country reports of progress on indicators in the Declaration of Commitment on HIV/AIDS adopted by the United Nations General Assembly in 2001.

For reproductive health in general, scores related to effort scores have been developed, though they have been applied in only four countries. These "Policy Environment Scores" focus heavily on the policy and organizational framework for subprograms under reproductive health, paying less attention to program components and service delivery (e.g., Almasarweh and Stover 1998).

In all these cases, effort or performance scores are ultimately based on subjective ratings. Given the complexity of the processes to be rated and variability across countries, comparable objective measures are unlikely to be available. Instead, reliance was placed on reputational measures. The situation has some parallels to that for corruption, for which objective measures are not readily available, particularly across countries. Early measures of corruption tended to be unreliable, being based on people's general impressions of the degree of corruption in a society. The weaknesses have been mitigated by carefully choosing respondents and designing questionnaires that focus on their actual experiences (Hawken and Munck 2007). This is the approach taken in this proposal.

The approach does have limitations. It is not meant to provide precise figures for progress on, for example, obtaining funding for a program. Results do not necessarily reflect official views. Questions may be difficult to match to specific national or international targets. However, these issues are addressed in the numerous other indicators already in use or suggested for the MDGs. Cross-checks with such objective measures may be possible, though only under a number of restrictive conditions: that the indicators have some arguable linkage, that they are available in the same time frame, that they cover enough units to be statistically relevant, etc. But any such comparisons obviously depend on having the performance scores available. The utility of the performance scores should be not in duplicating objective information but in providing a different perspective on the inevitably complex process of achieving the MDGs that is not tied only to what is already measured.

It is proposed that robust, uniform indicators be developed of each developing country's performance in achieving the MDGs, relying on questionnaires that produce expert ratings and on statistical procedures to provide indices from these ratings. This section discusses the questionnaires to be developed; the types of questions to be included and how the questionnaires are to be designed; how data will be collected; how the data will be screened, scored, weighted, and combined into indices; what reporting and dissemination options will be pursued; and what further data analysis should be planned.

Questionnaire form and content

Questionnaires are to be developed largely paralleling the lists of MDGs and the associated targets. These questionnaires are to be filled out by experts in particular sectors, those likely to be familiar with operational issues relating to achieving the targets. To facilitate a sector focus, some overarching goals have to be covered by several questionnaires. The goal of eradicating extreme poverty and hunger, for instance, requires two separate questionnaires, on performance (a) to improve income, employment, and equity and (b) to address issues of food and nutrition. Annex A provides a scheme for fitting particular goals and targets into a set of questionnaires.

To be convenient to answer and score, items in each questionnaire should almost all be structured, requiring yes/no answers or ratings on different scales. Some of the more complex issues may require brief open-ended answers, at least in a pilot stage, out of which structured questions can later be developed. Illustrative questions are shown in a sample questionnaire, for maternal mortality, in Annex B.

The content of the questions should vary, of course, by sector. However, two broad approaches to questioning can be expected. On one hand, questions should focus on the proximate determinants of the desired outcomes. These are understood as those features of services, or aspects of responses to them, or conditions that modify their effects, that most strongly and immediately determine service impact. On the other hand, questions could focus on the funding, infrastructure, facilities, training--the basic investments necessary to produce activity in the sector. The distinction somewhat parallels the earlier distinction between process and input, combined here under the rubric of performance or effort.

In the area of maternal health, for instance, access to emergency care in cases of pregnancy complications is essential to reducing deaths, given that these cases are often unpredictable. A proximate determinants approach prioritizes questions on such access: the degree of access in rural and urban settings; the availability of specific treatments, from oxytocin to Caesarian sections; the capacity of health facilities at different levels to provide them; the distance of facilities from clients; etc. An inputs approach complements this by emphasizing appropriate funding, construction of needed facilities, training of midwives, nurses, and doctors, etc.

Questions should be specific rather than general, asking for perceptions about aspects of sector operations that respondents are likely to have witnessed or at least have been aware of. Questions about overarching conditions that affect multiple sectors should be particularized. Governance, transport, and science and technology, for instance, each have broad impacts on multiple sectors. For the maternal mortality questionnaire, questions on transport could involve how pregnant women can access health facilities in emergencies. For a questionnaire on parasitic and infectious diseases, questions on governance could involve corruption in pharmaceutical procurement. For a food and nutrition questionnaire, questions on science and technology could involve agricultural research to improve strains or yields of major crops.

Distributional issues should be addressed in most questionnaires, with parallel or different questions for rural and urban areas, for regions, and for poorer as opposed to richer households. Rural and urban situations are often different and their services could be rated separately (where they are not covered in separate questionnaires, as they probably should be in the case of urban slums). In some cases, contrasts in goals for rural and urban areas, the required infrastructure, or the appropriate services may require that questions be framed differently. Water systems requirements are different if the intention is to support agriculture or to provide for urban needs. Environmental issues generally are different. Regional contrasts also need to be considered. Questions might include whether the regions most in need have been identified and how quality of services varies by region. Some regional variation may be difficult to get at with questions designed to be used cross-nationally, which may imply the need to use the questionnaire separately for each region (as was done for family planning effort in Vietnam [San et al. 1999] and for maternal health in India [Bulatao and Ross 2002]). Equity is a distributional issue that should not be confined to the questionnaire on income, employment, and equity but should find its way into most of the other questionnaires by way of an emphasis on services for the poor and the marginalized.

Gender issues are also distributional and require emphasis. Some specific gender targets have already been linked to particular questionnaires, and indicators of specific steps that might promote female welfare and empowerment (Corner and Repucci 2009) should appear not only in questionnaires directed to specifically female issues (such as maternal mortality) but also in other questionnaires.

Questionnaire development

Developing each questionnaire involves several steps.

- An initial, quick review of the literature, and where possible informal consultations with sector specialists, should lead to a list of possible topics and draft questions. As part of the review, data from similar previous questionnaires, if any, may require consideration and possibly reanalysis as a guide in questionnaire development.

- This should be followed by a small expert meeting,

for no more than half a day, to obtain comments, advice, and suggestions. Participants should be selected with convenience in mind. Others might be consulted through other means. This should lead to a substantially revised questionnaire.

- The questionnaire has to be pilot tested with at least a dozen respondents, to determine whether (a) it is comprehensible, easy to respond to, and of appropriate length; (b) it captures the essentials of performance in the area, from the perspective of the respondents; (c) the questions are applicable across diverse countries; and (d) different experts on the same country provide similar responses. This requires some respondents focusing on the same country and some focusing on different countries.

- Translation may be necessary into various languages. This should be limited if possible but may be unavoidable given the wide range of countries to be covered.

- Final versions of the questionnaire in the different languages to be used have to be produced both on paper and electronically.

Because of previous work in the area, a questionnaire on maternal mortality is one of the easier ones to design. Annex B provides such a questionnaire. Further consultations to refine it may be needed, and a pilot test is still essential.

Data collection and handling

Respondents for the final version of each of the questionnaires should be experts on and experienced observers of the relevant sector in the given country. In principle, all developing countries should be covered. Respondents could be drawn from the academic and research community, private organizations in the sector, international observers and advisers, and other pools of technical experts. Sector and government program managers could be included but should be identified as such, so that the possibility of biases can be checked. The number of respondents per questionnaire per country should be about 15-20.

To recruit these respondents, one of two approaches will be used in each particular setting. Where the sector experts are well known to UNDP local staff, these staff will take charge of selection and recruitment. Where staff are less familiar with the sector, a local consultant will be hired to be responsible for recruitment in consultation with UNDP staff, as well as for contacting respondents and collecting the data from them. Choosing the proper consultant for a given country is a key decision in ensuring adequate data.

Data from each questionnaire will require initial examination, particularly to determine whether, in a given country, respondents provide divergent ratings. Divergence and even wide disagreement in some cases can be expected and do not necessarily indicate invalid responses. They do however require examination, and some limited background information on the respondents should be collected to facilitate this.

Given that the questions relevant to one MDG may be in different questionnaires, data may have to be combined, for overall indices, across questionnaires. This would be facilitated if the data were collected more or less contemporaneously, to the extent this is possible, though for other reasons phasing the questionnaires would probably mean more attention to each questionnaire and better data. This and other issues relating to data management should be considered in questionnaire design.

Generating STEPS indices

Data from each questionnaire, which may contain 80-120 individual items, need to be combined into a smaller number of indices for each main topic covered, perhaps 20-30. An overall performance score for each country (or more than one overall score, if multiple targets are addressed in one questionnaire) is also needed. These indices could be produced a priori, based on the content of the questions and what aspects of performance the ratings are meant to reflect. However, statistical criteria can also be applied to produce robust indices.

A procedure for accomplishing this is outlined in Annex C. An overall index is assumed to be a linear combination of multiple ratings. A principal components analysis produces initial weights for all ratings, which are used in a Bayesian analysis to estimate final weights. A confidence interval for each resulting country index is automatically generated in the process.

The procedure is applied to earlier data on maternal mortality program effort (obtained with a questionnaire similar though not identical to that contained in Annex B). As Annex C shows, for these earlier data, the procedure gives a statistically robust index that provides a fair degree of statistical discrimination between countries.

Reporting, dissemination, and analysis

- Results of each investigation will be reported in several formats.

- A standard report will describe methodology and initial results. This should include all the estimated indices, and also possibly a classification of countries based on performance and progress toward the relevant MDG.

- A brief policy note will summarize this report and draw implications for action.

- Results will be incorporated into larger reports on progress toward the MDGs, especially the ten-year review.

- Country by country results will be posted on the Web using an online dashboard (illustrated in Annex D), a device that allows users to call up relevant data easily, readily visualizing the results and applying different weights to see how results could change.

- Briefings will be offered in particular countries where there is potential demand and particular policy implications.

The basic results should indicate performance relative

to the MDGs in each country, where effort is lacking, and whether the future trajectory looks positive or not for achieving the relevant MDG. Beyond these results, some of the worth of this exercise should lie in analysis to determine how performance in particular areas relates to progress toward the goals. The focus of national effort, in relation to the MDGs, is supposed to have been on using well tested approaches to produce relatively quick results, and one would therefore expect that effort indicators would show predictable links to outputs and outcomes. Demonstrating this with appropriate effort indicators still has considerable value, however, and one could always discover unexpected results or qualifications to the prevailing wisdom.

- Various types of analysis are possible with the cross-national data to be generated.

- Item analysis to attempt to refine performance

indices, to increase validity and reliability

- Analysis of the relationships among dimensions of effort to identify synergies in moving toward the MDGs

- Analysis of the links between dimensions of effort, on the one hand, and outputs and outcomes, on the other, to confirm the strategies recommended for moving toward the MDGs

- Analysis of how conditions outside a sector, such as the general quality of governance, impact on performance within the sector.

These analyses would give rise to a further set of reports. To be useful in accelerating progress toward the MDGs, this work would have to be done fairly quickly.

Table 1 indicates the time required for the investigation relating to a typical questionnaire. Roughly similar requirements are expected for each questionnaire, except for two: the maternal mortality questionnaire, where the drafting stage can be skipped, and whichever questionnaire comes after that, for which 10 percent more time should be allowed as staff become familiar with the details of producing effort indicators.

Some of the phases in this process may overlap, but the estimates are made excluding these overlaps. In particular, recruiting experts for a meeting and recruiting a consultant for each country are likely to require more time than is allowed and need to begin earlier, while a previous phase is in progress. In a later phase, data review need not wait for all the ratings to be completed.

One consultant is required per country (if this task cannot be handled by the UNDP Country Office), to be responsible for identifying respondents and obtaining their ratings. A lump sum payment per consultant of US$1,500 is suggested, possibly varying depending on country circumstances. Most of the rest of the work, including supervising the process, questionnaire drafting, and reporting, can be handled by one person on a half-time basis. Depending on the individual, short-term assistance may also be necessary in four respects with regard to organizing the expert meeting, translation, statistical analysis, and web site design both for the questionnaire and for reporting results. The meeting should be not more than a half-day meeting at UNDP headquarters, bringing in up to a dozen participants from relatively close by.

Whether investigations can be run in parallel depends on the staff resources to supervise them. Ideally, a couple of questionnaires should be completed first to iron out problems in the process. The remainder could probably be designed, fielded, and analyzed simultaneously, provided staff are available to supervise this.

Table 1. Approximate time required for a typical questionnaire

Activity	Duration in weeks
Literature review and initial questionnaire draft	4
Expert meeting, consultations, revised questionnaire	2
Pilot test, translation, final questionnaire	3
Recruiting consultants and respondents	9
Data review, reduction, summarization, report preparation	4
Dissemination activities	4
Further analysis and reports	8
Total	**34**

Abayomi K. 2009. Confidence intervals for linear combinations of dependent variables: MDG indexing. Unpublished paper. Industrial Systems and Engineering Department, Georgia Institute of Technology, Atlanta.

Adler N, Yazhemsky E, Tarverdyan R. 2009. A framework to measure the relative socio-economic performance of developing countries. *Socio-Economic Planning Sciences*, doi:10.1016/j.seps.2009.08.001.

Almasarweh I, Stover S. 1990. The policy environment score: Measuring the degree to which the policy environment in Jordan supports effective policies and programs for reproductive health. Washington, D.C.: Futures Group.

Berelson B. 1978. Prospects and programs for fertility reduction: What and where? *Population and Development Review* 4(4):579-616.

Bongaarts J. 1994. The impact of population policies: Comment [On Pritchett]. *Population and Development Review* 20:616-620.

Bulatao RA, Ross JA. 2001. Dimensions of ratings of maternal and neonatal health services: A factor analysis. Measure Evaluation working paper WP-00-39. Chapel Hill: University of North Carolina.

Bulatao RA, Ross JA. 2002. Rating maternal and neonatal health services in developing countries. *Bulletin of the World Health Organization* 80:721-727.

Bulatao RA, Ross JA. 2003. Which health services reduce maternal mortality? Evidence from ratings of maternal health services. *Tropical Medicine and International Health* 8:710-721.

Center for Global Development. 2009. Commitment to Development Index, Center for Global Development, Washington D.C. Accessed at http://www.cgdev.org/section/initiatives/_active/cdi/ on 20 October 2009.

CIESIN. 2009. Eco-region protection indicator for the natural resource management index of the Millennium Challenge Corporation. Accessed at http://sedac.ciesin.columbia.edu/es/mcc.html in May 2009.

Corner L, Repucci S. 2009. *A User's Guide to Measuring Gender-Sensitive Basic Service Delivery.* New York: UNDP and UNIFEM.

ESCAP. 2003. *Handbook on Reproductive Health Indicators.* New York: United Nations.

Freedman LP, Waldman RJ, de Pinho H, Wirth ME, Chowdhury AMR, Rosenfield A. 2005. *Who's Got the Power? Transforming Health Systems for Women and Children.* UN Millennium Project Task Force on Child Health and Maternal Health. London and Sterling, Va.: Earthscan.

Gelman A, Carlin JB, Stern HS, Rubin DB. 2004. *Bayesian Data Analysis*, 2nd ed. Boca Raton, Fla.: Chapman & Hall/CRC.

Hawken A, Munck GL. 2007. Measuring corruption: A critical assessment and a proposal. Technical Background Paper for Asia Pacific Human Development Report 2007. UNDP.

Hotez PJ, Molyneux DH, Fenwick A, Ottesen E, Ehrlich Sachs S, Sachs JD. 2006. Incorporating a Rapid-Impact Package for Neglected Tropical Diseases with Programs for HIV/AIDS, Tuberculosis, and Malaria. *PLoS Med* 3(5):e102.

Lapham RJ, Mauldin WP. 1972. National family planning programs: Review and evaluation. *Studies in Family Planning* 3(3):29-52.

Mauldin WP, Ross JA. 1994. Prospects and programs for fertility reduction, 1990-2015. Studies in Family Planning 25(2):77-95.

Mills S, Bos E, Lule E, Ramana GNV, Bulatao R. 2007. *Obstetric Care in Poor Settings in Ghana, India, and Kenya*. Washington, D.C.: World Bank.

OECD. 2005. Paris Declaration on Aid Effectiveness, High-level Forum on Joint Progress toward Enhanced Aid Effectiveness, 28 February to 2 March 2005, Paris, France. Accessed at http://www.oecd.org/dataoecd/11/41/34428351.pdf on 20 October 2009.

ONE. 2009. The Data Report 2009: Monitoring the G8 Promise to Africa, 19 May 2009. Accessed at http://www.one.org/international/datareport2009/pdfs/DR2009.pdf on 20 October 2009.

Policy Project. 2009. MNPI: Maternal and Neonatal Program Effort Index. Accessed at http://www.policyproject.com/pubs/mnpi.cfm in May 2009.

Pritchett LH. 1994. Desired fertility and the impact of population policies. *Population and Development Review* 20:1-55.

Ross JA, Stover J. 2000. New effort indices for family planning programs, 1999, and projections to 2004. Presented at the Conference on Family Planning Programs in the 21st Century, Dhaka, Bangladesh, January 17-20.

San PB, Ross JA, Phuong NL, Vinh ND. 1999. Measuring family planning program effort at the provincial level: A Vietnam application. *International Family Planning Perspectives* 25(1):4-9.

Stover J. 1999. The AIDS program effort index (API): Results from the field test. Washington, D.C.: Futures Group.

Stover J, Rehnstrom J, Schwartlander B. 2000. Measuring the level of effort in the national and international response to HIV/AIDS. Paper prepared for the 2000 International Conference on AIDS, July 9-14. Int Conf AIDS 13: abstract no. TuOrE336.

UN Millennium Project. 2005a. *Investing in Development: A Practical Plan to Achieve the Millennium Development Goals.* London and Sterling, Va.: Earthscan.

UN Millennium Project. 2005b. *Preparing National Strategies to Achieve the Millennium Development Goals: A Handbook.* New York: UNDP.

UNAIDS. 2004. 2004 Report on the Global AIDS Epidemic. Geneva: UNAIDS.

UNDP. 2009. UNDP's country analysis sheet (CAS), and its links to country action plans (CAP) for reaching the MDGs. Draft version, 16 April. New York: UNDP.

United Nations. 2003. *Report of the International Conference on Financing for Development, Monterrey, Mexico, 18-22 March 2002 (A/CONF.198/11, chapter 1, resolution 1, annex).* New York: UNDP.

United Nations. 2008. *The Millennium Development Goals Report 2008.* New York.

United Nations. 2009a. MDG Gap Task Force Report 2009. New York. Accessed at http://www.un.org/esa/policy/mdggap/mdg8report_engw.pdf on 20 October 2009.

United Nations. 2009b. *The Millennium Development Goals Report 2009.* New York.

USAID, UNAIDS, WHO, and the Policy Project. 2003. The Level of Effort in the National Response to HIV/AIDS: The AIDS Program Effort Index (API). Washington, D.C.: Futures Group.

A list of 14 questionnaires is suggested to parallel, though not exactly duplicate, the lists of MDGs and targets. These are listed in Table A1, which shows the goals and targets to which each refers. It also shows the output indicators related to each questionnaire that have been proposed in other documents. These output indicators were meant to be suggestive rather than comprehensive, presumably chosen at least partly for the availability of reliable data. What the questionnaires should address is the effort that has gone or is going into improving not only these outputs but also other outputs related to the broader goals and targets. The list in Table A1 follows the order of the MDGs. Some comments and qualifications on this list follow, in a different order, beginning with the social sectors, followed by economic issues, environmental issues, and the global partnership. (Further specification of what might be covered in each questionnaire is suggested in reviews of urgent issues in each sector, such as in UN Millennium Project 2005b:68-127.)

Five of the questionnaires are concerned with health issues. The family planning questionnaire already exists and current data are being collected. No additional effort is proposed in this area, except to incorporate results, as they become available, into overall assessments of effort toward achieving the MDGs.

A maternal and neonatal health questionnaire was also previously designed and used, as noted above, though the organizations responsible plan no further data collection. This questionnaire has been revised (Annex B) for potential use, though not pretested so far. Neonatal health is part of this questionnaire because of specific neonatal interventions that are integral to proper delivery care. This questionnaire should therefore provide some data about one set of interventions relevant to Target 4.A, reducing the under-five mortality rate. Results should be combined with data from the separate proposed questionnaire on child mortality to provide a proper picture of performance in this area. A similar approach to cross-cutting issues is to be used in other cases: questions should be addressed to those expected to be knowledgeable about particular issues, but the resulting data could be used in indices in a different area.

The child mortality questionnaire should cover at least immunization and integrated management of childhood illnesses. Ratings of effort on a neonatal package can be integrated with these topics to produce overall performance indicators. On HIV/AIDS, both prevention and treatment should be covered, as well as support for victims and their relations. In particular, education of orphans is to be covered under the education questionnaire. The last health questionnaire concerns other major infectious diseases, primarily malaria and tuberculosis. For particular countries, other tropical diseases may also deserve attention, such as schistosomiasis, hookworm, and ascariasis (Hotez et al. 2006). A cross-national set of questions on these diseases is unnecessary. Instead, this questionnaire should be devoted primarily to malaria and tuberculosis, with country-specific additions, where appropriate, for these other diseases. Access to essential drugs is also included in this questionnaire. It is a target under the global partnership, but this target also has operational implications for health systems that should be

considered in that context. Whether fewer questionnaires would be sufficient to cover the health sector is a relevant issue. Experience with the family planning and maternal mortality effort scores suggests that combining topics would not be useful. One needs to investigate operational issues in enough detail--which itself tends to make the results more reliable--and at the same time keep the questionnaires sufficiently short so that respondent cooperation does not become a problem. In fact, further disaggregation for the questionnaire on major infectious diseases might be considered at some point.

Also having a bearing on health are questionnaires on food and nutrition and on water and sanitation. These questionnaires must cover both the production side and the distribution and consumption side. For food and nutrition, improving agricultural and fisheries productivity must be considered, but also issues of food distribution, affordability, and consumption. Issues of micronutrient malnutrition might be important to address on a country-specific basis. Nutrition for mothers and infants needs attention, as does emergency food assistance.

For water and sanitation, the development of facilities should be considered, but also their actual use and upkeep. Urban and rural water systems involve different issues and technologies, requiring separate attention. Water and sanitation are covered in the MDGs as environmental issues, because their adequacy reflects the degree to which the environment is being conserved or not.

The education questionnaire should focus mainly on primary education, the second MDG. However, some attention is also needed to other issues: to secondary schooling, given the returns to it, particularly for girls; to higher education, given the need for advanced training to produce the staff for schools as well as for other interventions; to adult literacy, which is also a target; and to gender disparities at both primary and secondary levels, a target related to the gender goal; and to the schooling of HIV orphans, an indicator relative to HIV/AIDS.

The questionnaire on gender equity could focus on campaigns to eliminate violence against women, promote their property and inheritance rights, and empower them through grassroots initiatives such as microlending programs or broader attempts to achieve appropriate political representation. Developing performance indicators in this area should draw upon not only these data but also data from questionnaires on education, employment, maternal health, family planning, and possibly other areas.

The first MDG, to "eradicate extreme poverty and hunger," is partly addressed by the food and nutrition questionnaire. Specifying what should be covered under the remaining issues of income, employment, and equity is complicated. Three approaches, or some combination of them, are possible. First, one could recognize the importance of macroeconomic policy and planning, particularly whether policy is appropriate to particular country circumstances. One would have to go beyond the issue of official policy and raise questions about proper implementation, leading to issues of governance. Second, one could focus on rural and urban development separately. This could lead to some

overlap, since such issues as agricultural productivity, water and sanitation, and slum upgrading would be covered elsewhere. Third, the focus could be on specific government programs, such as income support programs, social protection programs, and special employment initiatives. Given the broad nature of the goal of poverty eradication, much of what might fall under this questionnaire should also be covered elsewhere, but this questionnaire should be conceptualized broadly so it is not just filling in gaps.

A separate questionnaire is meant to cover urban slums. This target is related to the goal of environmental sustainability, given that urban slums contribute greatly to urban environmental problems. The target, however, is not just to eliminate the negative environmental impacts but to improve the lives of slum dwellers. This may mean upgrading slum living in particular ways, such as regularizing tenure or improving transport, or facilitating the gradual movement of people out of slums. Low-cost housing on public land might be an option. One should at least consider what is being done to reverse the growth of slums, with measures to encourage people to stay in rural areas or reduce natural increase in urban areas.

A communication questionnaire is suggested by one of the

targets relating to the global partnership: to "make available the benefits of new technologies, especially information and communications." Much of the activity involves the private sector, though public involvement is required in policy, communications infrastructure, and connectivity for vital institutions.

The environment questionnaire, minus the water and sanitation and the urban slum parts, which are already covered, is still potentially very broad in coverage. Forests, freshwater resources, fisheries, air and water pollution, and greenhouse gases, and the protective measures required in each case, may need to be covered. Country by country the predominant issues may be different, but many countries may require at least some attention to each area.

For the global partnership for development, the issue is largely whether donor countries and agencies are performing as they should. Performance data tend to be more available on this issue than on others. Nevertheless, it is probably useful to obtain a bottom-up view of performance, i.e., ratings from the perspective of individual developing countries of how well the international community is performing in supporting development.

Table A1. Proposed questionnaires on effort and the goals, targets, and output indicators they should cover

Questionnaire	Covers these goals and targets	Covers activities to affect these output indicators (among other possible ones)
Income, employment, and equity	Goal 1: Eradicate extreme poverty and hunger Target 1.A: Halve, between 1990 and 2015, the proportion of people whose income is less than one dollar a day Target 1.B: Achieve full and productive employment and decent work for all, including women and young people	1.1 Proportion of population below $1 (PPP) per day 1.2 Poverty gap ratio 1.3 Share of poorest quintile in national consumption 1.4 Growth rate of GDP per person employed 1.5 Employment-to-population ratio 1.6 Proportion of employed people living below $1 (PPP) per day 1.7 Proportion of own-account and contributing family workers in total employment 3.2 Share of women in wage employment in the non-agricultural sector
Food and nutrition	Goal 1: Eradicate extreme poverty and hunger Target 1.C: Halve, between 1990 and 2015, the proportion of people who suffer from hunger	1.8 Prevalence of underweight children under five years of age 1.9 Proportion of population below minimum level of dietary energy consumption
Education	Goal 2: Achieve universal primary education Target 2.A: Ensure that, by 2015, children everywhere, boys and girls alike, will be able to complete a full course of primary schooling Target 3.A: Eliminate gender disparity in primary and secondary education, preferably by 2005, and in all levels of education no later than 2015	2.1 Net enrolment ratio in primary education 2.2 Proportion of pupils starting grade 1 who reach last grade of primary 2.3 Literacy rate of 15-24 year-olds, women and men 3.1 Ratios of girls to boys in primary, secondary and tertiary education 6.4 Ratio of school attendance of orphans to school attendance of non-orphans aged 10-14 years
Gender	Goal 3: Promote gender equality and empower women	3.3 Proportion of seats held by women in national parliament
Child mortality	Goal 4: Reduce child mortality Target 4.A: Reduce by two-thirds, between 1990 and 2015, the under-five mortality rate	4.1 Under-five mortality rate 4.2 Infant mortality rate 4.3 Proportion of 1-year-old children immunised against measles
Maternal and neonatal health	Goal 5: Improve maternal health Target 5.A: Reduce by three quarters, between 1990 and 2015, the maternal mortality ratio Target 4.A: Reduce by two-thirds, between 1990 and 2015, the under-five mortality rate	5.1 Maternal mortality ratio 5.2 Proportion of births attended by skilled health personnel 5.5 Antenatal care coverage (at least one visit and at least four visits) 4.2 Infant mortality rate
Family planning	Target 5.B: Achieve, by 2015, universal access to reproductive health	5.3 Contraceptive prevalence rate 5.4 Adolescent birth rate 5.6 Unmet need for family planning
HIV/AIDS	Goal 6: Combat HIV/AIDS, malaria and other diseases Target 6.A: Have halted by 2015 and begun to reverse the spread of HIV/AIDS Target 6.B: Achieve, by 2010, universal access to treatment for HIV/AIDS for all those who need it	6.1 HIV prevalence among population aged 15-24 years 6.2 Condom use at last high-risk sex 6.3 Proportion of population aged 15-24 years with comprehensive correct knowledge of HIV/AIDS 6.5 Proportion of population with advanced HIV infection with access to antiretroviral drugs
Major parasitic and infectious diseases	Goal 6: Combat HIV/AIDS, malaria and other diseases Target 6.C: Have halted by 2015 and begun to reverse the incidence of malaria and other major diseases Target 8.E: In cooperation with pharmaceutical companies, provide access to affordable essential drugs in developing countries	6.6 Incidence and death rates associated with malaria 6.7 Proportion of children under 5 sleeping under insecticide-treated bednets 6.8 Proportion of children under 5 with fever who are treated with appropriate anti-malarial drugs 6.9 Incidence, prevalence and death rates associated with tuberculosis 6.10 Proportion of tuberculosis cases detected and cured under directly observed treatment short course 8.13 Proportion of population with access to affordable essential drugs on a sustainable basis
Environment	Goal 7: Ensure environmental sustainability Target 7.A: Integrate the principles of sustainable development into country policies and programmes and reverse the loss of environmental resources Target 7.B: Reduce biodiversity loss, achieving, by 2010, a significant reduction in the rate of loss	7.1 Proportion of land area covered by forest 7.2 CO2 emissions, total, per capita and per $1 GDP (PPP) 7.3 Consumption of ozone-depleting substances 7.4 Proportion of fish stocks within safe biological limits 7.5 Proportion of total water resources used 7.6 Proportion of terrestrial and marine areas protected 7.7 Proportion of species threatened with extinction
Water and sanitation	Target 7.C: Halve, by 2015, the proportion of people without sustainable access to safe drinking water and basic sanitation	7.8 Proportion of population using an improved drinking water source 7.9 Proportion of population using an improved sanitation facility
Urban slum improvement	Target 7.D: By 2020, to have achieved a significant improvement in the lives of at least 100 million slum dwellers	7.10 Proportion of urban population living in slums
Communication	Target 8.F: In cooperation with the private sector, make available the benefits of new technologies, especially information and communications	8.14 Telephone lines per 100 population 8.15 Cellular subscribers per 100 population 8.16 Internet users per 100 population
Global partnership	Goal 8: Develop a global partnership for development Target 8.A: Develop further an open, rule-based, predictable, non-discriminatory trading and financial system. Includes a commitment to good governance, development and poverty reduction – both nationally and internationally Target 8.B: Address the special needs of the least developed countries, including: tariff and quota free access for the least developed countries' exports; enhanced programme of debt relief for heavily indebted poor countries (HIPC) and cancellation of official bilateral debt; and more generous ODA for countries committed to poverty reduction Target 8.C: Address the special needs of landlocked developing countries and small island developing States (through the Programme of Action for the Sustainable Development of Small Island Developing States and the outcome of the twenty-second special session of the General Assembly) Target 8.D: Deal comprehensively with the debt problems of developing countries through national and international measures in order to make debt sustainable in the long term	Official development assistance (ODA) 8.1 Net ODA, total and to the least developed countries, as percentage of OECD/DAC donors' gross national income 8.2 Proportion of total bilateral, sector-allocable ODA of OECD/DAC donors to basic social services (basic education, primary health care, nutrition, safe water and sanitation) 8.3 Proportion of bilateral official development assistance of OECD/DAC donors that is untied 8.4 ODA received in landlocked developing countries as a proportion of their gross national incomes 8.5 ODA received in small island developing States as a proportion of their gross national incomes Market access 8.6 Proportion of total developed country imports (by value and excluding arms) from developing countries and least developed countries admitted free of duty 8.7 Average tariffs imposed by developed countries on agricultural products and textiles and clothing from developing countries 8.8 Agricultural support estimate for OECD countries as a percentage of their gross domestic product 8.9 Proportion of ODA provided to help build trade capacity Debt sustainability 8.10 Total number of countries that have reached their HIPC decision points and number that have reached their HIPC completion points (cumulative) 8.11 Debt relief committed under HIPC and MDRI Initiatives 8.12 Debt service as a percentage of exports of goods and services

A questionnaire on effort at achieving the maternal mortality target is provided here. As noted in the text, it also covers neonatal interventions that should be associated with delivery, as well as family planning interventions that are appropriate postpartum.

Organization of the questionnaire

The questionnaire is organized in two parts. The first, much longer part requests ratings of different features of a maternal health program. The second, short part (labeled "General background") requests relatively objective information about laws, plans, budgets, facilities, etc. relating to maternal health. All respondents are expected to answer the first part, but only a few, those more closely connected with the government maternal health program, are to be given the second part to answer. Though the two parts are somewhat different in format, they are not separated so that respondents who receive both parts will see them as a single questionnaire.

Substantively, the questionnaire covers typical project components of policy and planning, funding, service delivery, and demand generation. However, questions are not posed in this order, but start with service delivery. The purpose is to fix the respondent's attention initially on what services actually reach women in need and can have direct effect on reducing maternal mortality. The questionnaire seeks to emphasize what is actually making a difference on the ground rather than what agreements and plans are made on paper. After asking about services in several different ways, the questionnaire moves to more general policy issues.

Questions are not necessarily grouped in categories familiar to donors. Instead, they are grouped for convenience, keeping together those with a similar frame of reference requiring answers in a similar format. Nor are questions intended as a checklist of all the specific requirements for providing proper maternal care. To keep the questionnaire at reasonable length and to avoid asking about details too fine for some respondents, the questions necessarily reflect

Table B1. Alternative classification of questionnaire items

Category	Questionnaire item numbers
1. Policy and planning: Policy is taken here in the sense of laws, regulations, standards, and guidelines that affect individual behavior relating to maternal health, the functioning of the maternal health program, and the conduct of service providers and others with whom they must interact. Plans are mainly national plans.	
1.1. Appropriate laws	59, 98, 99, 100a
1.2. Regulations and guidelines	56-60, 100
1.3. Plans	65, 71-72, 101-103
2. Budget and finance: Government budgets are covered as well as financing. Because cost recovery is discouraged in regard to maternal health services, local finance comes mainly from sources outside the health sector and is not specifically covered.	
2.1. Budget and expenditures	66-67, 69, 71, 74-76, 106-108
2.2. Donor support	68, 109, 112-116
2.3. Harmonization of activities	70, 110-111, 117
3. Service delivery: Different aspects of effective service delivery are listed below. Services usually require all of these elements to succeed, so most of the items could fall under most of the headings. However, each relevant questionnaire item is generally listed only under quality services and one other heading, reflecting its major emphases.	
3.1. Quality services	1-55
3.2. Adequate facilities	1-12, 20-21, 36, 72, 92, 95, 97, 105
3.3. Competent staffing	1-6, 10-18, 22-35, 49, 53, 78-86, 92, 104
3.4. Appropriate supplies and equipment	7-8, 47-48, 52, 73, 91
3.5. Equitable attention	9, 13-21,a 36-38, 74, 90, 65a, 65b, 86a, 97a
3.6. Effective monitoring and evaluation	60, 92-97
4. Demand generation: This involves mobilizing social groups and communities and providing good information to women and households about what needs to be done to avoid maternal deaths.	
4.1. Information, education, communication	62, 87-88, 91, 91a
4.2. Social mobilization	57, 89-90
5. Governance: Good governance cuts across the preceding categories requiring everything from sound policy to effective government services. One dimension of good governance as defined by the World Bank Institute, political stability and the absence of terrorism, is not directly assessed in its impact on the sector, and is left out here. A second dimension, government effectiveness, practically covers all the items, so it is represented by the more limited category of an effective management structure.	
5.1. Voice	57, 90, 97b, 97c
5.2. Effective management structure	61, 63-64
5.3. Regulatory quality	77
5.4. Rule of law	59b
5.5. Control of corruption	69, 76
a These items address rural-urban differentials.	
b This item suggests that the law should be disregarded in particular circumstances, in a sense going against the rule of law.	

a sampling of important best practices and dimensions of effort. To indicate how responses might be reclassified, after the data are obtained, to reflect particular issues of relevance from a planning perspective, Table B1 provides a possible alternative classification. The table lists some items more than once, as reflecting more than one aspect of performance. Some items could be listed under even more categories. Subsequent empirical analysis may suggest the most useful groupings.

Comparability of questionnaires

This questionnaire is a revision of one that has been used three times previously. Revising such a questionnaire requires difficult choices. Improvements are always possible, but too many changes reduce comparability with previous data and could make it difficult to rely on apparent trends.

This revision leans toward preserving comparability. No previous item was deleted. A number of new items are added, particularly in such areas as resources, policy, and delivery care. The second part of the questionnaire, on background information is also entirely new.

The additions were suggested by a review of program components and by such recent studies as a report on maternal health interventions (Freedman et al. 2005) and a set of case studies of maternal deaths and death rates (Mills et al. 2007). Plans for monitoring within World Bank reproductive health projects were also reviewed. UNDP staff and various reviewers suggested additional items.

The format of the original questionnaire has largely been maintained. However, retrospective ratings, asking respondents to rate not only current program effort but also effort as of three years previously, were dropped. These are no longer essential given earlier rounds of the questionnaire. Although the questionnaire has about 60 percent more items than previously (30 percent more if the second part is excluded), eliminating retrospective ratings should actually make it faster to complete.

Other format changes were possible but were not implemented. For example, all items were originally phrased positively, so that high ratings always represent strong effort. Phrasing some items negatively might have been useful to prevent respondents from developing a response set. Some items may be easier to respond to with percentages, rather than the true-false continuum mostly used. Some items might be convertible to simple yes/no items. None of these changes were made because the previous questionnaire appeared to be workable as is, and changes could impair comparability with previous data.

For the same reasons, the order of items has largely been kept as it was, though in adding items some rearrangement was needed. A few items were rephrased to improve readability, preserving the meaning and thrust of the originals.

Given that some comparability has been maintained with an earlier version, some further analysis of earlier data could give guidance on weights for indices and definition of confidence intervals (see Annex C).

QUESTIONNAIRE: MATERNAL AND NEONATAL PROGRAM EFFORT

COUNTRY: _____

RESPONDENT NAME: _____

ORGANIZATION: _____

POSITION / TITLE: _____

DATE: _____

NUMBER OF YEARS RESPONDENT WORKED:

	In Maternal Health	In Neonatal Health
At national level…………..	_____(years)	_____(years)
At provincial level………..	_____	_____
At district level……………	_____	_____
At community level………	_____	_____

THANK YOU FOR YOUR ASSISTANCE WITH THIS STUDY

INTRODUCTION

This questionnaire measures **maternal and neonatal program effort,** the effort that programs have put into reducing maternal mortality and morbidity and related neonatal risks. It focuses on strength of effort, meaning the volume and quality of program inputs and the adequacy of the process of providing services. It does **not** focus on outputs, such as cases served or maternal mortality rates. To keep the questionnaire short, only selected, representative functions are listed here.

Most items should be rated from 0 to 5, where 0 means the item is absent or extremely weak, and five means it is optimal. You can also think of each item as true or false; if it is entirely true it should receive a score of 5; if it is entirely false it should receive a score of 0. Intermediate situations should receive scores in between. A few items ask for your estimate of the percentage of the population with access to services.

The sections of the questionnaire cover these topics:

- *Staff at Health Facilities,* and what treatment they can provide for serious pregnancy complications

- *Access to Services,* in rural and urban areas

- *Antenatal, Delivery, and Neonatal Care,* and what components of each are or are not provided

- *Provision of Family Planning*, in connection with delivery care

- *General Supporting Functions,* including policy; resources; information, education, and communication; training; and monitoring, evaluation, and research.

Explanation of terms

Two types of facilities are referenced in this questionnaire. Facility types differ across countries, but generally the distinction is this:

HEALTH CENTER
Provides all essential and most emergency obstetric care. Midwives and nurses provide most of the care. Beds, anesthesia, sterile equipment, and supplies of drugs should be available.

FIRST REFERRAL FACILITY, OR DISTRICT HOSPITAL
Provides all levels of obstetric care, with trained staff including doctors who can perform Cesarean sections. Beds, anesthesia, sterile equipment, and supplies of drugs should be available.

Please indicate how true or false each statement is by circling a number, where 5 indicates "completely true," 0 indicates "completely false," and the numbers in between indicate partially true or false. (For example, a score of 2 on the first item below would mean that some health centers have staff who can handle manual removal of retained placenta, but somewhat more health centers do not have such staff.)

Staff at Health Facilities

		Completely false					Completely true
All Health Centers have trained staff, in place, who can provide these obstetric services:							
1.	Perform manual removal of retained placenta	0	1	2	3	4	5
2.	Perform vacuum aspiration of the uterus, using manual vacuum aspiration (MVA) or an electric suction device	0	1	2	3	4	5
3.	Use a partograph to determine when to refer	0	1	2	3	4	5
4.	Administer antibiotics intravenously	0	1	2	3	4	5
5.	Actively manage the third stage of labor—after the birth—with cord traction, uterine massage, and oxytocin	0	1	2	3	4	5
6.	Manage postpartum hemorrhage cases	0	1	2	3	4	5
7.	Have sufficient antibiotic supplies on hand of the correct types	0	1	2	3	4	5
8.	Can arrange transportation quickly to move a woman with obstructed labor to a district hospital	0	1	2	3	4	5
9.	Treat patients according to the same standards regardless of income or social background	0	1	2	3	4	5
All first referral facilities—District Hospitals—have trained staff, in place, who can:							
10.	Provide all functions listed above for Health Centers	0	1	2	3	4	5
11.	Perform blood transfusions, and have adequate supplies of safe blood on hand	0	1	2	3	4	5
12.	Perform Cesarean section or other operative delivery (e.g., forceps delivery or craniotomy)	0	1	2	3	4	5

Access to Services

Many women do not have access to a trained professional attendant, a health center with beds, or a district health center. Even if they do, the nearest attendant or facility may not have staff or equipment, or the service may be too expensive. This section pertains to the percentage of pregnant women with adequate access to each service. (Enter percentage on each blank.)

		Rural percentage			Urban percentage		
What percentage of pregnant women have adequate access to:							
13.	Treatment for postpartum hemorrhage during or soon after delivery						
14.	Management of obstructed labor						
15.	Treatment of abortion complications						
16.	Provision of safe abortion services, or menstrual regulation						
17.	Antenatal care during pregnancy						
18.	Delivery care by a trained professional attendant						
19.	Postpartum family planning services						
20.	District hospitals that are open 24 hours every day						
21.	Emergency transport to get to a health facility in an emergency						

Antenatal, delivery, and neonatal care

Please circle a number to indicate how true or false each statement is.

		Completely false					Completely true
At antenatal visits, all pregnant women:							
22.	Receive iron folate tablets for anemia	0	1	2	3	4	5
23.	Are examined for hypertension and treated as needed	0	1	2	3	4	5
24.	Are examined for syphilis and treated as needed	0	1	2	3	4	5
25.	Receive needed tetanus injections	0	1	2	3	4	5
26.	Are offered voluntary counseling and testing for HIV	0	1	2	3	4	5
27.	Are informed about danger signs in pregnancy and newborn complications	0	1	2	3	4	5
28.	Are assisted if necessary in preparing a written birth plan at least a month before expected delivery	0	1	2	3	4	5
At delivery (whether in a health facility or not), all pregnant women:							
29.	Are seen by a professionally trained attendant, whether at home or in a facility	0	1	2	3	4	5
30.	Have their labor monitored	0	1	2	3	4	5
31.	Are checked for signs of hypertension, anemia, and infection	0	1	2	3	4	5
32.	Are able to receive emergency obstetric care as needed	0	1	2	3	4	5
33.	Are provided an appointment for a check-up within 48 hours of delivery	0	1	2	3	4	5
34.	Are encouraged to immediately start breastfeeding their newborn	0	1	2	3	4	5
35.	Are counseled on umbilical cord care	0	1	2	3	4	5
36.	Attend a health facility if needed, and would not be deterred by the costs of doing so	0	1	2	3	4	5
37.	Can expect courteous and caring treatment wherever they deliver	0	1	2	3	4	5
38.	Can be confident that the cost of drugs, blood, and confinement, should an emergency develop, will not hamper treatment	0	1	2	3	4	5
		Completely false					Completely true
All newborns, whether delivered at home or in a facility:							
39.	Have their mouth and nasal passageways cleared	0	1	2	3	4	5
40.	Are dried and kept warm immediately after birth	0	1	2	3	4	5
41.	Receive prophylactic treatment for their eyes	0	1	2	3	4	5
42.	Have their umbilical cord cut with a clean blade	0	1	2	3	4	5
43.	Receive a DPT injection at three months	0	1	2	3	4	5
44.	Are scheduled for subsequent immunizations	0	1	2	3	4	5

Provision of Family Planning

		Completely false					Completely true
All Health Centers:							
45.	Routinely offer family planning after cases of abortion	0	1	2	3	4	5
46.	Routinely offer family planning at postpartum visits	0	1	2	3	4	5
47.	Have contraceptive pill supplies regularly in stock	0	1	2	3	4	5
48.	Have progestin-only pill supplies for breastfeeding women	0	1	2	3	4	5
49.	Have trained staff in place who can insert intrauterine devices	0	1	2	3	4	5
All first referral facilities, or District Hospitals:							
50.	Routinely offer family planning after cases of abortion	0	1	2	3	4	5
51.	Routinely offer family planning at postpartum visits	0	1	2	3	4	5
52.	Have contraceptive pill supplies regularly in stock	0	1	2	3	4	5
53.	Have trained staff in place who can insert intrauterine devices	0	1	2	3	4	5
54.	Can offer sterilization to female clients	0	1	2	3	4	5
55.	Can offer sterilization to male clients	0	1	2	3	4	5

Supporting Functions

Please circle a number to indicate how true or false each statement is.

		Completely false				Completely true	

Policy

		Completely false					Completely true
56.	Ministry of Health policies toward pregnancy and delivery services are adequate.	0	1	2	3	4	5
57.	Policies are developed through adequate consultation with interested parties, such as other ministries, NGOs, private practitioners, and women's groups.	0	1	2	3	4	5
58.	Policies are reasonable and fair concerning which personnel can provide maternal health services (e.g., allowing trained midwives to perform a wide range of medical procedures).	0	1	2	3	4	5
59.	Policies require immediate treatment of complications of abortions, regardless of its legality.	0	1	2	3	4	5
60.	Policies relating to maternal health are vigorously implemented through regular high-level reviews and updated action plans.	0	1	2	3	4	5
61.	The manager for maternal health programs is placed at a high administrative level.	0	1	2	3	4	5
62.	High government officials, including those in the Ministry of Health, issue frequent statements to the press and public supporting improvements for safe pregnancy and delivery.	0	1	2	3	4	5
63.	The maternal and child health program is a respected component of health activities and is more often admired than belittled, ignored, or resented within the Ministry of Health.	0	1	2	3	4	5
64.	District health system managers have the authority they need to direct effective maternal and child programs in their districts.	0	1	2	3	4	5
65.	Government has a poverty reduction strategy that includes reducing maternal mortality as a key element.	0	1	2	3	4	5
65a.	Maternal health services are available equally to all women regardless of marital status, race, ethnicity, religion, politics, or income.	0	1	2	3	4	5
65b.	Reproductive health services and practices are required to be respectful of the culture and practices of all minorities.	0	1	2	3	4	5

Resources

		Completely false					Completely true
66.	The government budget for safe pregnancy, delivery, and postpartum care (for facilities, personnel, supplies, etc.) is adequate, whether the funds are from the Ministry of Health, provincial government, or donor support.	0	1	2	3	4	5
67.	The maternal and child health program can expect a regular flow of funds to cover its current activities as well as to improve its facilities.	0	1	2	3	4	5
68.	Donor support is appropriate to the needs of the maternal and child health program and is provided in timely fashion.	0	1	2	3	4	5
69.	The entire amounts allocated (whether adequate or not) are regularly utilized for maternal health services, with little or no wastage or leakage of funds.	0	1	2	3	4	5
70.	Other health initiatives do not distract from and do not take resources away from maternal health programs.	0	1	2	3	4	5
71.	Resources for maternal health services are allocated to regions and districts annually based on a review of actual needs.	0	1	2	3	4	5
72.	Emergency obstetric facilities are adequate in number and properly located, or the government has concrete plans to build enough of them.	0	1	2	3	4	5
73.	All health centers have emergency power supplies in case of electrical outages.	0	1	2	3	4	5
74.	All maternal health services and drugs are provided free to all clients.	0	1	2	3	4	5
75.	Neither pregnant women nor their families ever have to provide the drugs or medical supplies they need during confinement.	0	1	2	3	4	5
76.	Maternal health staff do not solicit money or gifts beyond approved fees.	0	1	2	3	4	5
77.	The private sector—doctors, midwives, clinics, maternity homes—is active and handles a substantial share of pregnancy and delivery cases.	0	1	2	3	4	5
78.	Doctors and nurses in maternal health tend to stay in government service rather than pursuing other options.	0	1	2	3	4	5
79.	Vacant positions for doctors and nurses in maternal and child health are filled without excessive delay.	0	1	2	3	4	5
80.	Doctors and nurses in maternal health are usually at their posts during scheduled hours rather than being absent.	0	1	2	3	4	5

		Completely false				Completely true	

Training

81.	Medical curricula include hands-on clinical training in obstetric care, including management of several deliveries.	0	1	2	3	4	5
82.	All midwives and nurses in health centers have received refresher training for safe pregnancy and delivery care within the last five years.	0	1	2	3	4	5
83.	Doctors in health centers have received refresher training for safe pregnancy and delivery care within the last five years.	0	1	2	3	4	5
84.	Midwives and nurses recently hired for health centers receive training in safe pregnancy and delivery care within the first six months.	0	1	2	3	4	5
85.	Newly hired doctors receive special in-service training for normal deliveries.	0	1	2	3	4	5
86.	Enough midwives are being trained for local needs.	0	1	2	3	4	5
86a.	All maternal health care providers are required to be proficient in the languages spoken in their assigned areas.	0	1	2	3	4	5

Information, education, communication

87.	The national program uses the mass media to educate the public about symptoms of pregnancy complications and safe places for childbirth.	0	1	2	3	4	5
88.	The national program also uses the mass media to educate the public about harmful home practices in pregnancy care, delivery, and postpartum care (home remedies and birthing customs, etc.).	0	1	2	3	4	5
89.	Community organizations take part in programs to educate the public about safe pregnancy and delivery.	0	1	2	3	4	5
90.	Women play the major role in campaigns to reduce maternal deaths.	0	1	2	3	4	5
91.	The Ministry of Health supplies adequate educational materials (posters, pamphlets, etc.) to delivery facilities to instruct clients about safe practices.	0	1	2	3	4	5
91a.	A comprehensive sex education program is provided for in-school and out-of-school youth, covering among other things unwanted pregnancy and HIV/AIDS.	0	1	2	3	4	5

		Completely false				Completely true	

Monitoring, evaluation, research

92.	A routine statistical system uses facility reports to provide good periodic information on supplies, personnel, deliveries, Cesarean sections, and cases of complications.	0	1	2	3	4	5
93.	Staff at the national level regularly monitor and analyze results from the facility reports.	0	1	2	3	4	5
94.	Recent surveys provide data on maternal events (pregnancies, deliveries, attendants and sites for deliveries, estimates of maternal deaths, etc.).	0	1	2	3	4	5
95.	An updated listing exists of facilities that can treat obstetric emergencies.	0	1	2	3	4	5
96.	Administrators in the Ministry of Health systematically use statistical information for decision making and periodically reconsider strategies for reducing maternal mortality.	0	1	2	3	4	5
97.	Each hospital follows a regular procedure to review and learn from each case of a maternal death in the facility.	0	1	2	3	4	5
97a.	All data collected on maternal mortality indicate the race or ethnicity of the woman.	0	1	2	3	4	5
97b.	Each maternal care client is informed about how to file a complaint, if desired, about the quality, efficiency, and equal provision of maternal and neonatal health services.	0	1	2	3	4	5
97c.	Each complaint filed is taken up by a designated authority, and the complainant receives a response.	0	1	2	3	4	5

General Information

The following questions provide background information useful in understanding the context of the reproductive health program. Please fill in the blanks or check the appropriate alternative.

Plans and regulations

98.	Women are not legally allowed to marry until age:	_____ years

99.	Abortion (whether or not it is easily available) is legal:	
	with few if any restrictions	○ 5 ○ 4
	in a limited set of circumstances	○ 3 ○ 2
	in no circumstances	○ 1 ○ 0

100.	Are there locally adapted guidelines for antenatal care?	○ Yes	○ No
100a.	Do the nation's laws recognize a universal right to reproductive health care?	○ Yes	○ No
101.	Does the government have a current Medium Term Expenditure Framework (MTEF)?	○ Yes	○ No
102.	Is funding for activities to reduce maternal mortality explicitly incorporated into this framework?	○ Yes	○ No
103.	Does the government have a Poverty Reduction Strategy (PRS) that includes specific goals and indicators for reducing maternal mortality?	○ Yes	○ No

Inputs

104.	How many doctors, nurses, and midwives are there in government service? Doctors _____ Nurses _____ Midwives _____		
105.	How many health centers are there in urban and rural areas? Urban health centers _____ Rural health centers _____		
106.	Of the annual government budget, what percent is allocated for health?		_____ percent
107.	Are there official figures on how much the government spends on health services annually per capita?	○ Yes	○ No
108.	If so, how much is this (including donor funds if any)?		

Donors		

109.	Please check each of the donors that have been active in the health sector in the last three years, and check also which have contributed to the reproductive health program in this period.	

	Active in health sector	Active in reproductive health
International and multilateral organizations		
WHO	❍ 1	❍ 2
UNICEF	❍ 1	❍ 2
UNFPA	❍ 1	❍ 2
UNDP	❍ 1	❍ 2
Global Fund to Fight AIDS, Tuberculosis, and Malaria	❍ 1	❍ 2
World Bank	❍ 1	❍ 2
African / Asian / Inter-American Development Bank	❍ 1	❍ 2
Others (please list):		
	❍ 1	❍ 2
	❍ 1	❍ 2
	❍ 1	❍ 2
National development aid agencies		
Australian Agency for International Development (AusAID)	❍ 1	❍ 2
Canadian International Development Agency (CIDA)	❍ 1	❍ 2
Danish International Development Agency (DANIDA)	❍ 1	❍ 2
Gesellschaft für Technische Zusammenarbeit (GTZ)	❍ 1	❍ 2
EuropeAid	❍ 1	❍ 2
French Development Agency (AfD)	❍ 1	❍ 2
Japan International Cooperation Agency (JICA)	❍ 1	❍ 2
Millennium Challenge Corporation (U.S.)	❍ 1	❍ 2
Norwegian Agency for Development Cooperation (NORAD)	❍ 1	❍ 2
U.K. Dept. for International Development (DFID)	❍ 1	❍ 2
U.S. Agency for International Development (USAID)	❍ 1	❍ 2
Others (please list):		
	❍ 1	❍ 2
	❍ 1	❍ 2
	❍ 1	❍ 2
	❍ 1	❍ 2
Non-governmental organizations		
Aga Khan Development Network	❍ 1	❍ 2
Oxfam International	❍ 1	❍ 2
Project Hope	❍ 1	❍ 2
Others (please list):		
	❍ 1	❍ 2
	❍ 1	❍ 2
	❍ 1	❍ 2
	❍ 1	❍ 2
	❍ 1	❍ 2

110.	In the last three years, Ministry staff in reproductive health have met with:	
	All active donors all together at least once	❍ 5 ❍ 4
	The majority of active donors, though not all together	❍ 3 ❍ 2
	Few if any of the active donors	❍ 1 ❍ 0

111.	Coordination among donors in reproductive health is:	
	Excellent	⚪5 ⚪4
	Average	⚪3 ⚪2
	Poor	⚪1 ⚪0
112.	Donor funding in reproductive health in the last three years has been:	
	Generous and sufficient for the needs	⚪5 ⚪4
	Helpful though short of reasonable expectations	⚪3 ⚪2
	Well short of the need	⚪1 ⚪0
113.	In providing support for reproductive health, donors:	
	Usually insist on their own ideas	⚪5 ⚪4
	Push some ideas but also accept local initiatives	⚪3 ⚪2
	Allow local Ministry staff to determine the needs	⚪1 ⚪0
	Usually improved the program	⚪5 ⚪4
	Sometimes improved the program	⚪3 ⚪2
	Rarely improved the program and sometimes hurt it	⚪1 ⚪0
115.	Technical assistance provided by donors for the reproductive health program:	
	Has made it more effective	⚪5 ⚪4
	Has had little effect	⚪3 ⚪2
	Has wasted resources and distracted staff	⚪1 ⚪0
116.	In evaluating donor funding for reproductive health, one could say it has:	
	Filled critical needs	⚪5 ⚪4
	Helped in some ways but not others	⚪3 ⚪2
	Made little difference or been detrimental	⚪1 ⚪0
117.	Donor support for health services other than reproductive health has:	
	Generally improved the reproductive health program	⚪5 ⚪4
	Neither helped nor hurt the reproductive health program	⚪3 ⚪2
	Generally weakened the reproductive health program	⚪1 ⚪0

Introduction

The goal is to generate a country-level index of performance relative to a particular MDG target or goal. This index should be a linear combination of possibly nonindependent variables. This section investigates the methodologies to build the weighting scheme and define confidence intervals for such an index.

The data

The data arrive as a collection of ratings or scores X_i, on questionnaire items X_1 through X_k, with some multivariate, non-independent, distribution f_X.

$$X = (X_1, \ldots, X_k) \sim f_X$$

Each X_j is an "average" from judges (say $1, \ldots, N$) and the Confidence Interval (CI) for country i should be of the form

$$P\left(\theta_i \in (L_i, U_i)\right) = 1 - \alpha$$

where (Li,Ui) are the lower and upper bounds for country i.

A useful cross country index from such data would include a confidence interval (CI) for each country. Let the value of the index for country i be θ_i, a weighted sum of judge ratings X_1 through X_k. That is:

$$\theta_i = \sum_{j=1}^{k} c_j X_j$$

The vector c^T is the weighting scheme chosen for the index. Under the assumption that this scheme is constant across countries $i = 1, \ldots, N$, the CI's for each country i should be a function of the particular choice of the scheme c^T as well as of distributional assumptions regarding X.

Let $\mu^T = (\mu_1, \ldots, \mu_k)$ be the vector of means for the variables X in the index and let $\sigma^{2^T} = (\sigma_1^2, \ldots, \sigma_k^2)$ be the vector of variances. Note that $\sigma_{j,l} \equiv Cov(X_j, X_l)$ and collect the variances and covariances in the covariance matrix. The correlation is $\rho_{j,l} = \frac{\sigma_{j,l}}{\sigma_j \bullet \sigma_l}$; collect the correlations as $\rho^T = \left(\left(\rho_{j,l}\right)\right)_{j<l=1\ldots K}$.

Confidence Intervals

Three possible methods of generating country-wide confidence intervals are:

- Distribution Free: minimal assumptions are placed on multivariate distribution of the Judges' ratings.

- Frequentist: Distributional assumptions on f_X, the multivariate distribution of X.

- Bayesian: Prior distributions on the parameterization of f_X.

These approaches are listed in order of the restrictiveness of a priori assumptions: distribution free (distribution invariant) approaches impose the least assumptions on the data – the

Bayesian approaches require more supposition. Generally, a more definite (assumptive) model yields tighter confidence intervals for the parameter estimates.

We used what could be called a 'simple' or 'naïve' Bayesian approach in this example – in that the prior distributions are assumed conditionally independent. This is a commonly used approach on many types of data. Illustrations of the distribution-free and frequentist approaches are outlined in the supporting document Abayomi (2009) listed in the references. Gelman et al. (2004) is a good reference for the Bayesian approach to data modeling.

Bayesian Approach

The Bayesian approach is to add an additional distribution on the parameters of interest: we introduce probability distributions for the country score θ and propose probability distributions for the parameters μ^T and Σ, the means and covariances of the judge's ratings.

Multivariate Normal

Σ known: Consider the case when the covariance matrix for X is known or well estimated. The prior distribution

$$\mu^T \sim N(\mu^T_0; \Lambda_0)$$

assumes that the means are multivariate normal with $\mu^T_0; \Lambda_0$ fixed (i.e. estimated from data). The *posterior distribution* for μ^T is then

$$\pi\left(\mu^T \mid x, \Sigma\right) \equiv N\left(\mu_0^T, \Lambda_0\right)$$

where

$$\mu_n = \left(\Lambda_0^{-1} + n\Sigma^{-1}\right)^{-1} \left(\Lambda_0^{-1}\mu_0 + n\Sigma^{-1}\overline{x}\right)$$

and

$$\Lambda_n^{-1} = \Lambda_0^{-1} + n\Sigma^{-1}$$

Here x are the n observed judge ratings. Note that θ is merely a linear transform of x, in vector notation: $\theta = c^T X$. Thus θ is univariate normal with

$$E(\theta) = c^T\mu_n$$

and

$$Var(\theta) = Var(c^T\mu_n) = c^T\Lambda_n c$$

Bayesian CI's (often called Credible Intervals) are the random draws from this distribution. The posterior distribution here being multivariate normal. In this case we have closed form expressions for the expectation and variance of θ. A reasonable approximate Bayesian CI is

$$(1-\alpha) \equiv P\left(\theta \in c^T\mu_n \pm Z_{\alpha/2} \bullet c^T\Lambda_n c\right)$$

Σ unknown: The results are similar with the additional relaxation of a prior on the variance- covariance matrix Σ as well. A common prior is:

$$\Sigma \sim Inv-Wishart_{v0}\left(\Lambda_0^{-1}\right)$$

and

$$\mu \mid \Sigma \sim N\left(\mu_0, \Sigma / \kappa_0\right)$$

where v_0 and κ_0 are the degrees of freedom and scale matrix for the inverse-Wishart distribution on Σ. The joint posterior is multivariate normal. Sampling from the joint posterior to generate CI's for θ can follow this algorithm:

Draw $\Sigma \mid x \sim Inv-Wishart_{v0+n}\left(\Lambda_n^{-1}\right)$

Draw $\mu^T \mid \Sigma, x \sim N\left(\mu_n, \Sigma / \kappa_n\right)$

Compute $\theta = \mathbf{c}^T\mu$

This yields a sampling posterior for θ and the CI can be gleaned directly from inspection of the simulated replicates.

Weighting

Choosing the appropriate weighting scheme and generating CI's for each scalar θ_i are separable tasks. The CI's are of course affected by the choice of weighting scheme, however, the weights themselves are arbitrary in the sense that they are subject to an exogenous constraint chosen by the indexers.

Desirable conditions on the choices on the weights could be:

- Maximal independence within X

- Minimum covariance between X_i and X_j

- Maximum variation across scores θ_i

Maximal Independence

Consider a model

$\mathbf{Y} = B\mathbf{X}$

where the components of Y are independent, and B is an estimate of A^{-1}, a mixing matrix for the latent/unobserved model:

$\mathbf{X} = A\mathbf{S}$

with $S \sim Q = \prod_{i=1}^{K} Q_i$. This is the Independent Component Analysis (ICA) model and algorithms exist to estimate B and thus the θ as $\hat{S}$.

Consider a diagonalization of B

$B = \mathbf{L}^T D\mathbf{L}$

with L an upper triangular matrix, and D a diagonal matrix. D yields a weighting scheme for the components of X and could be used as weights cT. Alternately, since $Y_j = B_j\mathbf{X}$ - the `independent' output of the ICA algorithm could be used as proxies for X in a null weighting scheme.

Minimum covariance

Principal Component Analysis (PCA) can be viewed as a special case of the above ICA approach where Q is a multivariate Gaussian distribution. The diagonalization of B is immediately

$B = \Delta^T\mathbf{E}\Delta$

where Δ and E are the eigenvectors and eigenvalues of the covariance matrix Σ in Q. Weighting items or components in this scheme is essentially factor analysis.

Maximum variation across scores

The output of the MDG indexing - a presentation of country-by-country scores (with confidence intervals and ranks) - suggests that maximizing variation across scores (across countries) is a desirable feature of a weighting scheme.

This goal may be addressed in a repeated measurement extension of the ICA or PCA algorithms, where the individual judge ratings are collected over all countries $\mathbf{X}_i = 1 \ldots N$.

Bayesian Weighting

A direct approach is to let the $\mathbf{c}^T$ weights themselves have a prior distribution and investigate the distribution of θ with this additional prioritization.

This is to model θ as univariate normal as above:

$$\theta \sim N\left(c^T\mu_n, c^T\Lambda_n c\right)$$

with

$\mu^T \sim N(\mu^T_0; \Lambda_0)$

and

$$\Sigma \sim Inv-Wishart_{v0}\left(\Lambda_0^{-1}\right)$$

and

$c^T \sim Dirichlet(\alpha)$

Sampling from the joint posterior to generate CI's for θ can follow this algorithm:

Draw $c^T \mid \mathbf{x} \sim Dirichlet(\alpha)$

Draw $\Sigma \mid x \sim Inv-Wishart_{v0+n}\left(\Lambda_n^{-1}\right)$

Draw $\mu^T \mid \Sigma, x \sim N\left(\mu_n, \Sigma / \kappa_n\right)$

Compute $\theta = \mathbf{c}^T\mu$

with $\alpha_1 = \ldots = \alpha_k = 1$; μ_n; κ_n and Λ_n^{-1} as before. In a Monte Carlo procedure this program is iterative and repeated until tolerance limits on the distribution of the parameters are satisfied.

An Exploration Building on a Previous Questionnaire

In 1999, a survey for maternal health with a similar structure to the one being designed for the MDG STEP Index was carried out in several countries (Bulatao and Ross 2002). The data contained in this survey provides an opportunity for testing the methodology for the robustness of the index. The

methodology should be able to tell us:

- The main issues that drive performance at a country level (i.e. discriminate the main drivers of variability, hence the weighting scheme needs to be appropriate and the same across countries).

- Allow us to discriminate across countries (i.e. the methodology should be able to determine statistically significant index levels across countries).

The survey provides us with $N = 1037$ observations by $K = 182$ variables: the judge ratings with metadata. The metadata are country and judge specific information.

The rating data are variables 21-101- variables 102-182 are repeated measurements by each judge. These are the judge scores - **X** - as outlined in section 1.1, above.

X are the judge ratings, variables 21-101, and the metadata are variables 1-20 including country name and id.

Data Preparation

The entire data (including the repeated measurements) have 9505 missing values; 319 of the missing values are in the metadata for the judges.

The percent of missing items is low (5 percent) but non-negligible. The location of the missing data, however, cannot be ignored. Missing data in both the meta-data and the covariates are imputed via *hot-deck,* this is, the completed data are re-samples of the observed at each country. A feature of the *hot-deck* procedure is that the model for the completions is explicitly empirical. The data were completed by *hot-deck* at each country to avoid collecting error beyond each set of country rankings.

The observations for Tanzania were discarded as many covariates were completely missing for all judges, thus reducing the total data to N = 1022.

PCA for null weighting

Recall that the goal is to generate a score at each country which is a linear combination of the judge's ratings, $\theta_i = \sum_j c_j X_{\cdot j}$. A priori, without any index or response variable to calibrate an initial or null weighting, a decision rule for the scheme can be minimal variance across rating items.

This procedure provides a projection of the collected rating items, the variables, to an orthogonal or independent basis. Weights assigned via a minimal variance scheme can identify (Gaussian or Normal) overdetermination in the covariates and suggest which may be discarded or of redundant importance in an index. See Bulatao and Ross (2001) for a prior application of factor analysis to these data.

Results and Discussion

Aggregating variation across Judges

The PCA procedure, highlighted above, is used to generate a set of null weights c. An initial PCA on the ungrouped data suggests the presence of some redundancy in the covariates; 28 percent of (Gaussian) variation can be explained by only one component, out of 81 possible.

The elements of the first eigenvector for the PCA decomposition are used as null weights:

each $c_j \equiv e_j \Big/ \sum_j e_j$. Thus each cj (0, 1) and Σj cj = 1.

This approach generates an index score for each judge, (and therefore several for each country). Following this method, the maximum score was a judge rating for Gujarat and the minimum score was for a rating in Yemen.

Null weighting by PCA when aggregated across judges may introduce inordinate bias to account for the variation within country, across judges. Notice that the maximum index score was generated by one (perhaps) optimistic rater for Gujarat.

Aggregating variation across Countries

Following the PCA procedure, but this time aggregating across countries estimates the eigenvectors, the null weights, via decomposition of the covariance matrix on the countries, instead of on the judges. This aggregation explains a higher proportion of the variation in the ratings, as can be seen in Figure 1. Following this process the maximum score was for Jamaica and the minimum for Yemen.

Bayesian weighting

The scores generated by the PCA weighting are used as initial values in a Bayesian method for estimating the weights.

This is the scheme:

Generate c_0^T as elements of first eigenvector from PCA. These null weights yield $\theta_0 = c_0^T$ **X**, the null scores.

Generate Var(θ_i) = c_0^T Var(**X**)**c**, the variance within a judges rating.

Estimate Var(θ_0) as the sample variance of the null scores.

The PCA procedure provides the initial scores θ_0 (generated from the null weighting scheme) and estimates for between and across variance.

- Let $\theta_i \sim N(\beta_g, \sigma_i)$, where the initial value of $\sigma_i = \sqrt{Var(\theta_i)}$. Here i = 1, . . . ,N, the number of judges.

- Let $\beta_g \sim N(\mathbf{c}^T\mathbf{X}, \sigma_g)$ be the country scores, where the initial value of σ_g is set to $\sqrt{Var(\theta_0)}$.

- Let $c_j \sim$ *Dirichlet* (α) be the distribution for the weights. The initial weights are set identically to 1.

This scheme allows a posterior to be estimated for σ_g and c_j; the country specific scores and the variable weights. The posterior distributions yield confidence intervals for the country scores and the associated weights, automatically.

If all the judges ratings come from distributions with equivalent support - like {1,2,3,4,5} for Likert type or [0,1] for percentages, say - the values of the weights can be interpreted as relative importance. The value of the weight for each item is the contribution of the item to the overall score, with respect to the way in which the weights are estimated.

In the example, the initial weights are assigned to maximize discrimination among countries; the resulting estimates are the relative contributions of items under this paradigm. These initial weights are starting estimates for the joint conditional estimation of the scores, weights, and associated variation.

Choosing a different weighting paradigm, via an alternate scheme, such as maximum variation among groups of countries or maximum inner product or score; (see Adler et al. 2009 for an alternative approach) yields different relative importances, of course, but with the same interpretation - modulo the method. Of course, the weighting scheme may be adjusted to reconcile the judges' responses, especially when the questions have nonequivalent support, such as some being "yes/no" items and others being rated {1,2,3,4,5}. The adjustment should leave the interpretation of the estimated weights unchanged.

Summary statistics generated from the posterior distributions are included with this report. Plots of the posterior distributions of the parameters for the country scores and variable weights are in figures 2 and 3.
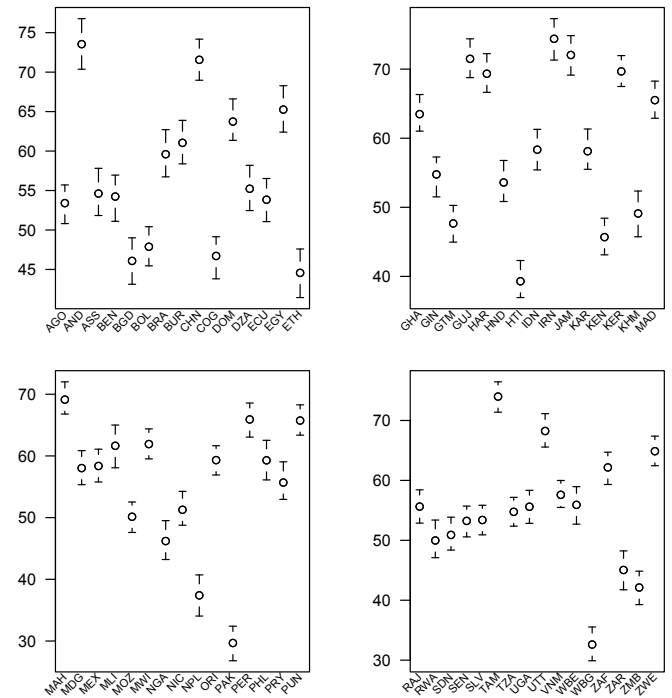
Figure 1: Screen plots for variation of PCA by component. The left graph is the variation explained across judges, the right is across countries. A first component explains, respectively, 28 and 41 percent of the variation for each aggregation
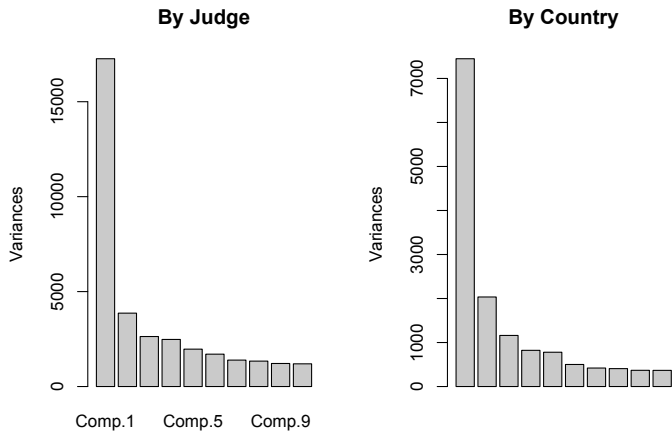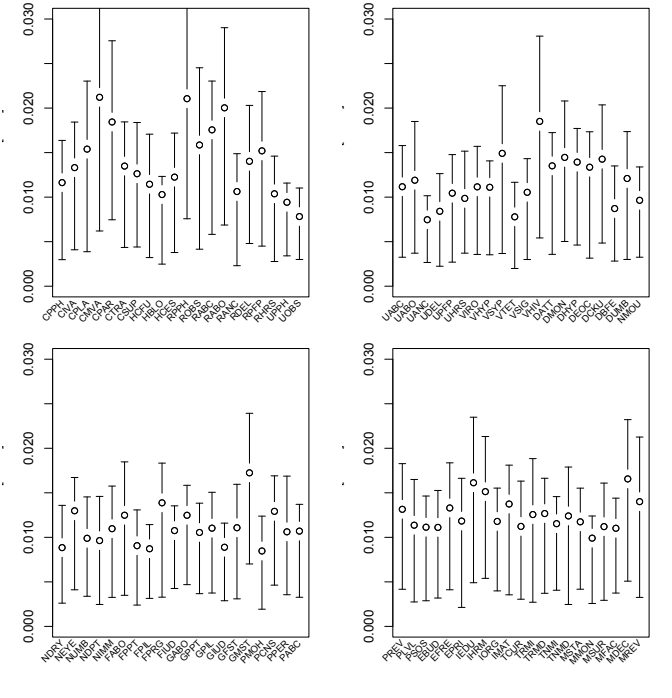

Figure 2 shows that the CIs obtained using the Bayesian methodology satisfy all the goals set for the index at the onset, allowing for clear discrimination across countries while obtaining robust statistically significant results.

Figure 2: Distribution of country scores, from posterior replicates, by alphabetical order of ISO3 country id code. The upper and lower `whiskers' are the 75th and 25th percentiles of the posterior distribution.

Figure 3: Distribution of variable weights, from posterior replicates, by order of variable in questionnaire. The upper and lower `whiskers' are the 75th and 25th percentiles of the posterior distribution.

An online dashboard is a Web-based platform to collect, organize, and display data in a standardized, simplified manner. It can be used, for instance, to display country data on education (as in Figure D1, which however uses different data from those to be collected under this proposal), facilitating the input of new data and comparisons across countries and over time.

It is proposed that a dashboard be used in collating responses on the various questionnaires covered in this proposal. The dashboard should then allow the calculation of various indices from the response data, such as indices on performance relative to policy and planning, budget and finance, service delivery, demand generation, and sector governance, as well as on overall sector effort. Users--including UNDP, governments, and other stakeholders--can then resort to the dashboard to track performance, toggling between different data sets and time periods (once the data have been entered) to obtain snapshots of progress in each country and pinpoint bottlenecks in progress toward particular MDGs.

Figure D1. An illustrative use of an online dashboard for education data