

**Measurement and Indicators for  
Impact Evaluations of Supply-Side  
Results-Based Financing (RBF) for Health  
Programs**

**Prepared by: Shivam Gupta and David Peters  
Johns Hopkins University Bloomberg School of Public Health  
2010**

**Revised by:  
Christel Vermeersch, World Bank  
2012**

## Table of Contents

1. Goal.....	3
2. Indicator selection criteria .....	3
3. Sources of information for impact evaluation .....	5
4. Supply side RBF evaluation framework .....	6
5. Proposed indicators .....	8
6. Measuring quality of health services .....	9
7. References.....	13
Annex 1: Sources of indicators.....	15

## 1. Goal

The goal of this paper and its companion analysis sheet is to:

- Summarize, organize and classify the information that is being collected using the model HRITF questionnaires
- Propose standard indicators and ways of calculating those indicators, so the information extracted from different evaluations is comparable
- Help principal investigators understand what information is included in the HRITF model instruments, as well as what is not

The document is meant to be a living document that should be updated when the HRITF instruments are piloted in country and analysis is completed on project.

## 2. Indicator selection criteria

The selection of proposed indicators for the RBF project was guided by a set of criteria tied to the overall goal of contributing to global learning (Bryce, Victora et al. 2010). For example, each individual indicator should be:

- **Valid & Reliable** - It must be an indicator that measures the phenomenon in question, and provides data that is reproducible across settings and over time. This implies standardization of the definition used for each indicator (i.e., numerators and denominators, where appropriate) as well as when and how it is measured.
- **Relevant, important and likely to change as a result of inputs** - The indicator must address an important part of the pathway from health systems strengthening through service delivery, quality and coverage to impact that is being addressed by enough of the projects to make cross-site comparisons feasible. Being central to this pathway also implies that the indicator will show significant change during the project period as a result of project inputs.
- **Feasible for measurement** - The indicator should be able to be measured with resources available through the project or through collaboration with other activities planned or under way in each setting, with more than one measurement occurring within the time the project is active. Some factors to be considered are costs related to travel, supervision and the infrastructure required to collect, store and analyze the data for indicators.

- **Consistent with global standards** - Standard indicators that have been defined and are already being used by other initiatives and projects should be adopted whenever possible, to allow comparisons and promote global learning.

The set of indicators should be:

- **Limited in number** - Not all possible indicators can be included or the data bases will be unwieldy and unnecessarily duplicative with the data sets developed and maintained by other projects. In addition, the attention span and time of households and health workers is limited, so it's best to keep questionnaires as short as possible.
- **Amenable to linked hierarchical analysis** - The units of analysis for the indicators should vary across the topical areas to be addressed by the project. For example, individuals are the most appropriate unit for measurements of population health, but health facilities may be the most appropriate unit for measurements of service quality given the strategies to be implemented. The set of core indicators will need to be organized in ways that permit the linking of measurements across these various types of units, referred to as hierarchical analysis.
- **Focused and yet reasonably comprehensive relative to project outputs and outcomes** - innovative approaches like RBF pose a particular evaluation challenge in that it is not always clear ex ante what the channels of impact will be. Investigators will need to balance the need to be focused (and not let the scope of data collection run out of hand) with the need to be comprehensive in picking up changes in intermediate indicators including output indicators.

The indicators listed in this paper include custom indicators for RBF, as well as a number of indicators that are common with other sources including the Millennium Development Goals (MDG's), the WHO World Health Statistics 2011 Compendium, the Countdown to 2015, the WHO toolkit to measure health system strengthening, the health facility assessment and readiness to provide services assessments developed by Measure evaluation, and the Maternal and Child Health (MCH) program documentation guidelines for the Catalytic Initiative. More details on those sources can be found in Annex 1.

The instruments included in the Impact Evaluation (IE) toolkit are meant to be a general guide to investigators trying to determine the content of their data collection tools. They are sufficiently flexible that modules or units can be left out or added depending on the research questions of the particular project.

### 3. Sources of information for impact evaluation

The IE toolkit includes several data collection instruments for supply-side RBF programs.

- A household survey instrument
- A health facility survey instrument

While the household survey instrument can also be used for demand-side RBF programs, it does not include specific modules on the demand-side program, which would need to be custom-developed for each evaluation.

**Household survey:** The household survey instrument is a population-based tool that aims to generate data on representative populations or sub-populations (usually defined in terms of geographic, socio-economic or gender characteristics). Household surveys are important for impact evaluation because they relate to the whole target population, and not only to groups using institutional services. At the same time, household surveys are periodic, that is, they are not continuously generated but rather applied at a certain point in time, as opposed to administrative data.

**Health facility surveys:** These provide another important methodology for collecting data on health services and for validating routine health service data, by observing service delivery, inspecting facilities, interviewing staff and clients, and reviewing archives. A health facility survey involves visiting and collecting data from a representative sample of all health facilities in a country or geographical area. Alternatively, a health facility census involves collecting such data from all health facilities in a country or geographical area. Such a survey/census may consider different aspects of service quality, such as the availability of drugs, commodities and trained staff. Special techniques, such as record review, observing client-provider interaction and using standardized patients (sometimes called mystery clients), add considerable value to the assessment. However they also increase the costs and complexity. Data collected from record reviews and staffing inventories can be used to validate routine administrative statistics on the volume of services delivered and on the availability and geographical distribution of human resources.

In addition to household surveys and health facility surveys, impact evaluations can benefit from having additional sources of data which are not specifically collected for the purpose of the evaluation.

**Census:** The population and housing census is the primary information source for determining the size of a population and its geographical distribution, plus the social, demographic and economic characteristics of its people. It can provide vital statistical data

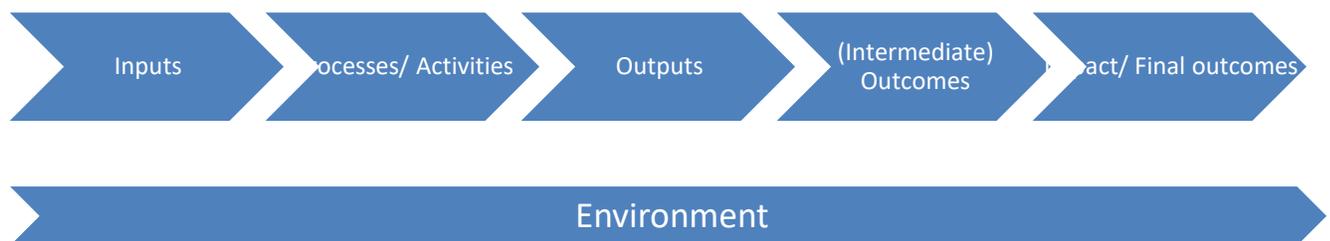
on population and housing situations at even the smallest administrative levels. From the health perspective, information on population numbers and distributions by age, sex and other characteristics is essential for national and local planning, estimating target population sizes and trends, and evaluating service-coverage rates and future needs. Information on major determinants and risk factors, such as poverty, housing conditions, water supply and sanitary facilities may also be included. The disadvantage, however, is that only a small number of health questions can be included.

#### 4. Supply side RBF evaluation framework

The basic framework for development of indicators for Impact Evaluation of Supply side RBF programs is a results chain. There are many different versions of results chains that can vary in their use of terminology. It is important to remember that results chains are a way to help conceptualize the transformation of inputs into final outcomes, and one should not get bogged down too much into different authors' interpretation of what are the precise boundaries between outputs, outcomes, and final outcomes. A practical way to look at it is to think of those boundaries as a convention, and once one has agreed on the convention to use, it is simply a question of being consistent.

Below, we use a modified version of the results chain framework proposed for evaluation of the scale-up for better health in the context of the IHP+.<sup>1</sup> The top section shows the results chain sequence from inputs and activities/processes to outputs, outcomes and impact. These five domains are systematically linked to each other. Each domain results from other domains placed to its left, for example, project inputs and processes/activities lead to project outputs. Key contextual factors that may affect progress in the results chain are included at the bottom of the figure.

The RBF results chain used here and throughout the toolkit is a generic one that will need to be adapted before it can be used for any particular country. This results chain is not meant to be prescriptive, but rather it is used in a generic manner to identify a number of core indicators that can be measured as part of evaluating the RBF projects.



---

<sup>1</sup> (Bryce, Victora et al. 2010)

**Inputs** - Domestic and international inputs are listed in the leftmost side of the figure. These include input-based funding, in-kind inputs, performance-based funding, technical assistance, domestic and foreign expertise, human resources, etc.

**Processes /Activities-** Processes include the “machinery” put in place in the context of RBF, including the processes for defining, billing, paying and verifying service delivery.

**Outputs** -The middle box of the framework shows the expected outputs of introducing the results-based financing mechanism. Outputs span material resources (buildings, equipment, drugs, medical products) in as far as they are affected by RBF, provider knowledge in as far as it can be affected by RBF, provider motivation and effort, organizational structure of facilities, existence and use of information technology, accessibility of care and the process of clinical and interpersonal care.

**Outcomes** - The increased outputs resulting from Results-based financing are expected to translate into improved outcomes, including overall utilization and equity of utilization, and quality of care provided to patients. Utilization is defined as the proportion of the population who receive an intervention among those eligible for it. Utilization is influenced by supply (provision of services), demand and uptake by people in need of services.

**Impact** - Ultimately, the increased coverage is expected to improve health, modulated by the efficacy of the interventions. For maternal, neonatal and child health interventions, the primary impact should be reductions in child and maternal mortality and morbidity, and improvements in nutritional status.

**Environment** - Environmental factors are those factors that affect the impact of RBF interventions. They are the modulating factors and covariates of the impact regressions. Some clear environmental factors are the political climate, the socio-economic status of the target population, or the geographic isolation of communities and health facilities. Other environmental factors are not so clearly categorized, in the sense that they can also be *affected by* RBF, such as the stock of health workers, the knowledge of health workers, or the organizational structure of health facilities.

- Inputs
  - Funding
  - Human Resources (number and qualifications)
- Activities/Processes
  - Definition of RBF activities
  - Billing
  - Payment
  - Verification
- Outputs
  - Availability of trained health workers \*

- Material resources (buildings, equipment, medical products drugs) \*
- Provider knowledge \*
- Provider motivation and effort \*
- Organizational structure of facilities \*
- Existence and use of information technology
- Process of clinical care \*
- Process of interpersonal care \*
- Outcomes
  - Utilization of services \*
  - Coverage of services \*
  - Equity of service utilization \*
  - Equity of service coverage \*
- Impact
  - Health outcomes for children (0-5) \*
  - Health outcomes for women (15-49) \*
  - Health outcomes for other layers of the population (\*)
  - Equity of health outcomes \*

## 5. Proposed indicators

In the companion Excel file, we propose two sets of core and optional indicators, that could be used for the RBF projects according to the domains of the results chain: Input, Activities/Processes, Output, Outcome and Impact. Information on the numerator, denominator and the data collection method is provided for each of the proposed indicators.

## 6. Measuring quality of health services

The proposed results chain framework contributes to better understanding and measurement of efforts to evaluate maternal and child improvement programs in particular RBF programs. However, it is important to point out that the measurement of quality of care in health systems does not fit neatly into one of the categories of the results chain, but instead it spans across multiple categories.

According to the widely accepted model proposed by **Donabedian**, quality of care can be measured across three domains - **structure, process and outcome** (Donabedian 1980; Donabedian 1988). **Structure** relates to the characteristics of the system in which care is delivered: it includes attributes of material resources (building, equipment, availability of services, examinations and drugs), human resources (number and qualification of personnel, knowledge) and organizational structure (medical staff organization). Structural features of health care provide the opportunity for individuals to receive care but do not guarantee it, although they can have direct impact on processes and outcomes. **Process** is the actual delivery or receipt of health care. Process of care in a health system has been described under two categories: clinical care and interpersonal care. Clinical care refers to the application of clinical medicine to a particular health problem that a particular individual is suffering from, whereas interpersonal care describes the interaction of health service provider and the treated individual. **Outcome** measures the impact of care on the health status of the users. This is measured in terms of the change in health status (functional status, clinical outcome) and user satisfaction.

Two particular dimensions of quality of care deserve further elaboration because their measurement is not straightforward and rapidly evolving: **provider knowledge and delivery of health care**. In the results chain, provider knowledge can be both an input to the RBF program but also an output if RBF affects knowledge; delivery of care spans outputs and outcomes. In the Donabedian framework, provider knowledge is part of structural quality, while delivery of care is part of the process.

A number of methods have been used to measure provider knowledge and delivery of health care (Franco, Daly et al. 1997; Hermida, Nicholas et al. 1999; Peabody, Luck et al. 2000; Bessinger and Bertrand 2001; Franco, Franco et al. 2002; Leonard and Masatu 2005; Leonard and Masatu 2006; Das and Hammer 2007). Table 4 provides a summary of the quality of care methods, by comparing their validity, reliability, feasibility and relative costs.

Sheet 2 of the companion Excel file provides the detailed list of indicators and indices that can be pulled from the HRITF questionnaires for measuring the quality of care provided at primary health care facilities, with a focus on maternal and child health. For each indicator and index, we note whether it is applicable to (1) Sick children under five years of age; (2) children under five visiting the facility for immunization and routine well child visits; (3) antenatal care for pregnant women, and (4) post natal care for women after delivery. The indicators to measure quality of care for these services overlap to a certain extent, but there are also differences in equipments, drugs and clinical protocols used for each service.

The impact evaluation toolkit contains three types of vignettes:

- **Case scenario vignettes (kindly shared by Johns Hopkins University) – page 12 of the health worker questionnaire**

In these vignettes, the enumerator reads the case of a patient with particular symptoms, and asks the provider for all of the actions and prescriptions that the provider would take to provide this patient with the most appropriate treatment. In most of the scenarios, apart from the initial reading of the case, the enumerator does not provide any further information to the provider. However, in one of the vignettes, the case has two parts.

- **Vignette on protocol for prenatal care – page 13 of the health worker questionnaire**

This vignette tests provider knowledge of the protocol for a first prenatal visit for a pregnant woman. The case scenario does not contain any information on additional symptoms other than visible pregnancy.

- **World Bank-ISERDD vignettes (kindly shared by Jishnu Das and team)**

In these vignettes, the enumerator reads an initial case of a patient with particular symptoms. The enumerator then asks the provider to treat him/her as if he/she were the patient. The provider can ask the enumerator/“patient” any question to be able to come to a diagnosis, and can also propose exams. The enumerator/“patient” provides the answer to the questions of the provider, and gives the results of the proposed exams to the provider. The provider is asked to come to a diagnosis and treatment proposal. There may be follow-up questions from the enumerator as to the treatment that the provider would give to the patient. *Please note that these vignettes are included in a different file in the toolkit.*

**Table 4: Summary of methods to assess provider knowledge and delivery of care.**

Method	Strengths	Weaknesses	Reliability	Feasibility	Relative Costs
<b>Provider knowledge</b>					
Provider interviews – knowledge, attitudes	Measures practical knowledge; measures provider perceptions	Hawthorne effect – Altered behavior or performance of the health worker resulting from awareness of being part of an experimental study; Does not measure actual practice	Good for provider perceptions and knowledge (but many poor questions are used)	Can be done in large sample sizes, wide variety of cases, including rare events	Low
Provider vignettes	Measures practical knowledge, decision-making	Hawthorne effect – Altered behavior or performance of the health worker resulting from awareness of being part of an experimental study; Does not measure actual practice	Variability based on interviewer; can have low reliability (though may be good in certain circumstances)	Depends on high quality interviewer, uses lower sample size and limited variation in cases, some rare events	Higher interviewer qualifications and training required
<b>Health service delivery</b>					
Observation of care	Measures actual quality delivered	Hawthorne effect – Altered behavior or performance of the health worker resulting from awareness of being part of an experimental study; Measures optimum care delivered	Good -- Limited inter-rater problems, but not well documented	Requires high case load &/or common conditions; more intrusive than other methods	Modest training (e.g. 1 week) of literate or student health workers; supervision costs
Exit Interview – History taking and physical exam	Lower Hawthorne effect than observation if provider can be blinded to study	Does not measure actual sequence of physical exam	Poor patient/caretaker recall: low correlations with actual history documented	Can do large numbers, but better for common conditions	Modest training (e.g. 1 week) of literate or student health workers; supervision costs

<b>Method</b>	<b>Strengths</b>	<b>Weaknesses</b>	<b>Reliability</b>	<b>Feasibility</b>	<b>Relative Costs</b>
Exit interview – Counseling, Perceptions, Patient characteristics	Measures what caretaker actually understands; Measures immediate perceptions ; Can measure key characteristics (e.g. wealth) which can be compared with population parameters	Perceptions at point of care may not reflect those measured later	Good – limited inter-rater problems; some questions have shown good correlation with actual history	Can do large numbers if facilities have high volumes, better for common conditions	Modest training (e.g. 1 week) of literate or student health workers; supervision costs
Simulated clients (mystery patients)	Lowest Hawthorne effect if able to keep interviewer blinded	Poor verisimilitude for children and women in delivery  Not possible for pregnant women and sick children (ethical issues)	Variability based on actor	Low sample size; least variation in conditions to test	Higher interviewer qualifications and training can increase costs
Patient record review	If good record keeping, can reflect care intended to be provided; Better for diagnosis and treatment (if systematically recorded)	Poor records are the norm; Limited information on actual tasks performed reflects what providers say patient care was rather than actual care delivered;	Poor, especially for non-standardized record keeping	Non-intrusive, quick, but records rarely of adequate quality for use other than volume of service	Cheapest

## 7. References

- Bessinger, R. E. and J. T. Bertrand (2001). "Monitoring quality of care in family planning programs: A comparison of observations and client exit interviews. ." International Family Planning Perspectives **27**(2): 63-70.
- Bryce, J., B. Daelmans, et al. (2008). "Countdown to 2015 for maternal, newborn, and child survival: the 2008 report on tracking coverage of interventions." Lancet **371**(9620): 1247-1258.
- Bryce, J., C. G. Victora, et al. (2010). Evaluating the scale-up for maternal and child health. A common framework. Under review.
- Das, J. and J. Hammer (2007). "Location, location, location: residence, wealth, and the quality of medical care in Delhi, India." Health Affairs **26**(3): w338-w351.
- Donabedian, A. (1980). The definitions of quality and approaches to its assessment. Ann Arbor, Health Administration Press, School of Public Health, The University of Michigan.
- Donabedian, A. (1988). "The quality of care. How can it be assessed?" JAMA **260**: 1743-8.
- Franco, L. M., C. C. Daly, et al. (1997). "Quality of case management of sexually transmitted diseases: comparison of the methods for assessing the performance of providers." Bulletin of World Health Organization **75**(6): 523-32.
- Franco, L. M., C. Franco, et al. (2002 ). "Methods for assessing quality of provider performance in developing countries." International Journal of Quality in Health Care. **14**(17-24).
- Gilroy, K., E. Hazel, et al. (2010). Guidelines for documenting program implementation and contextual factors in independent evaluations of the Catalytic Initiative, Institute for International Programs, Johns Hopkins Bloomberg School of Public Health.
- Hermida, J., D. D. Nicholas, et al. (1999). "Comparative validity of three methods for assessment of the quality of primary health care." International Journal of Quality in Health Care. **11**(5): 429-33.
- HMN (2008a). Framework and Standards for country health information systems. Health Metrics Network. Geneva, World Health Organization.
- IHP+ (2008). Monitoring performance and evaluation of the scale for better health. A proposed common framework, International Health Partnership+.
- Leonard, K. L. and N. C. Masatu (2005). "The use of direct clinician information and vignettes for health service quality evaluation in developing countries." Social Sciences and Medicine **61**(9): 1944-51.

- Leonard, K. L. and N. C. Masatu (2006). "Outpatient process quality evaluation and the Hawthorne effect." *Social Sciences and Medicine* **63**(9): 2330-40.
- Measure Evaluation (2007). Guidance for selecting and using core indicators for cross country comparisons of HF readiness to provide services. Health Facility Assessment Technical Working Group.
- Measure Evaluation (2008). Profiles of Health Facility Assessment Methods. Report of the International Health Facility Assessment Network (IHFAN).
- OECD (2006). "Survey on monitoring the Paris declaration." *OECD Journal of Development* **8**(2).
- Peabody, J. W., J. Luck, et al. (2000). "Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality." *JAMA* **283**(13): 1715-22.
- UNICEF (2008). Tracking progress in maternal, newborn and child survival: The 2008 report. New York, United Nations Children's Fund.
- United Nations (2009). The Millennium Development Goals Report 2009.
- Walford, V. (2007). Indicators on health systems performance, Department for International Development.
- WHO (2008). Measuring Health System Strengthening and trends. A toolkit for countries. Draft document, World Health Organization.
- World Bank (2008). Global impact evaluation program for the health results innovation grant. Draft concept note. , The World Bank.
- World Bank (2009). Monitoring performance-based funding or results-based financing schemes. First meeting of the RBF monitoring informal interest group. Washington DC, The World Bank.

## Annex 1: Sources of indicators

- **Millennium Development Goals (MDGs):** The Millennium Development Goals (MDGs) are eight international development goals that all 192 United Nations member states and at least 23 international organizations have agreed to achieve by the year 2015. They include reducing extreme poverty, reducing child mortality rates, fighting disease epidemics such as AIDS, and developing a global partnership for development. MDGs 4 and 5 specifically deal with reduction of child and maternal mortality respectively (United Nations 2009).
- **WHO World Health Statistics 2011- Indicator Compendium**
- **Countdown to 2015:** The Countdown to 2015 collaboration aims to track coverage for interventions that are essential to attainment of Millennium Development Goals (MDG) 4 and 5 and some elements of MDGs 1, 6, and 7. Countdown to 2015 represents a common evaluation framework for tracking coverage of proven interventions and measures of mortality and nutrition in countries with the highest burden of mortality in mothers and children. Since its inception in 2005, the Countdown process has evolved by shifting its focus from attention to the individual child or mother to focus on the continuum of care from before pregnancy, through to pregnancy, childbirth, and the postnatal period, and on to early childhood. Tracking of single, biologically based interventions was complemented by inclusion of broader approaches or packages, such as antenatal or postnatal care, that can serve as platforms for delivery of multiple interventions (Bryce, Daelmans et al. 2008).
- **WHO toolkit to measure health system strengthening:** Since 2000, WHO has led an initiative to reach broad-based consensus around key indicators and effective methods and measures of health systems capacity, including inputs, processes and outputs, and to relate these to indicators of outcome. This toolkit proposes a core set of indicators and related measurement strategies that can assist countries, global health initiatives and other stakeholders to monitor and direct investments into health systems (WHO 2008).
- **Health facility assessment and readiness to provide services by MEASURE evaluation:** The Health Facility Assessment technical working group (HFA – TWG) funded by United States Agency for International Development (USAID) identified the difference in definition of indicators and data elements as a key gap in various facility-based methods of health system monitoring and evaluation. These differences limit comparisons of data across surveys, countries, and time. The TWG proposed a core set of indicators to accelerate the

availability of data useful for health system analysis at a minimum cost while reducing redundancies and duplication of data-collection efforts. These indicators are defined and collected the same way, regardless of the approach used for collecting the facility-based data (Measure Evaluation 2007; Measure Evaluation 2008).

- **MCH program documentation guidelines for the Catalytic Initiative:** The Catalytic Initiative to Save a Million Lives (CI) is a central part of the movement to accelerate progress toward MDG4, and includes major efforts by UNICEF, CIDA, AusAid, the Bill and Melinda Gates Foundation to support efforts by countries and their partners, including the United Nations. All CI partners are committed to evaluating the success of the initiative in accelerating reductions in under-five mortality. Documentation of program implementation and the contextual factors that may affect child survival are essential components of any evaluation effort. The purpose of these guidelines is to help governments, their partners, and in-country research partners develop and implement plans for this component of the evaluation. The guidelines provide a generic framework to clearly document and describe how and what level of quality the activities are implemented in different settings (Gilroy, Hazel et al. 2010).