A Review of Best Practices in Performance Reporting: Developing a Broad Hierarchy of Performance Measures

Report: July 15, 2012

Prepared by:

Onil Bhattacharyya, MD, PhD
Anita McGahan, MBA, PhD
Will Mitchell, PhD
Kate Mossman, PhD
Raman Sohal, MA
John Ginther, BA
Jieun Cha, MSc
Ameya Bopardikar, BHSc
John A. MacDonald, MD
Leigh Hayden, PhD
Himanshu Parikh, MD
Ilan Shahin, MD, MBA
Table of Contents

Executive Summary .................................................................................................................. 2
Introduction ............................................................................................................................... 3
Method ..................................................................................................................................... 3
    Collecting Supplementary Information from Publicly Available Sources ................................ 3
Reporting Practices among Well-Documented Programs .......................................................... 6
Overview: Evaluating Indicators .............................................................................................. 8
    Indicators ............................................................................................................................ 8
    Targets/Benchmarks ............................................................................................................ 8
    Annualizing Data ............................................................................................................... 8
    Ranking Indicators in the CHMI Plus Framework .................................................................. 9
Measuring Population Coverage ............................................................................................ 10
Measuring Health Output ........................................................................................................ 11
Measuring Health Outcome and Impact .................................................................................. 14
Measuring Affordability ......................................................................................................... 17
Measuring Physical Access and Service Availability / Allocation ............................................ 19
Measuring Pro-Poor Targeting ............................................................................................... 21
Measuring Clinical Quality ..................................................................................................... 25
Measuring User Satisfaction ................................................................................................... 27
Measuring Management Quality ............................................................................................ 29
Measuring Human Resources Supply and Commitment of Staff ........................................... 33
Measuring Political Support .................................................................................................... 36
Discussion ............................................................................................................................... 40
    Credibility, Feasibility, and Comparability ......................................................................... 41
    Strengths ............................................................................................................................ 42
    Limitations ......................................................................................................................... 42
Conclusion and Next Steps ...................................................................................................... 43
Executive Summary

The growing number of organizations using market-based approaches to address the health care needs of the poor in sustainable and scalable ways has generated great interest among governments and private investors. The Center for Health Market Innovations (CHMI) has developed a database of innovative programs that seek to provide accessible quality care in low and middle-income countries. This database can help identify best practices for replication and expansion by governments, other organizations, and a range of potential investors. To be useful, it requires comprehensive information that is credible and allows for comparison within and across strategies and disease areas. The health services literature has several frameworks for performance measurement, but it is unclear whether these approaches are feasible in resource-constrained settings. This report reconciles established frameworks from the academic literature with what is reported by programs listed in the CHMI database to describe best practices in performance reporting. This work will provide guidance to funders and those who evaluate programs as well as program managers on a comprehensive framework to develop indicators that are credible, feasible, and comparable.

Reporting Practices of Well-Documented Programs

In a previous report, we found that the CHMI dataset had few comprehensive descriptions of programs, so we supplemented this with publicly available data on 80 programs, selecting from top-reporting programs, mHealth, and maternal and child health. This data was used to revise the CHMI performance measurement framework, producing 12 descriptive categories and 19 indicator categories. Adding publicly available data on these 80 programs to what was in the CHMI database increased the proportion of fields with information from 30% to 55%. The most frequently reported fields were health output (86.25%), operational and technical partners (82.5%), affordability (71.25%), and balance sheet (68.75%). The number of indicators for each field varied according to its complexity, with health output, health outcome, and management quality reporting the highest median number of indicators.

Best Practice in Performance Measurement

We reviewed the 80 programs across 11 indicator categories and selected the data most closely aligned with accepted definitions from the health services literature. Ranking criteria were used to develop good, better, and best indicators for these 11 fields, as well as recommended operational definitions for each category. These definitions were based on the optimal balance of credibility, feasibility, and comparability, and each dimension of performance was rated on these characteristics. Credibility is most relevant for investors and researchers, and the most credible indicators were in financial management, health output, and population coverage. Feasibility is especially relevant for program managers who are responsible for data collection, and the most feasible indicators were in health output, human resources supply/commitment of staff, and management quality. Comparability is important to researchers and investors that are interested in benchmarking, but comparability was only high for financial management and moderate to low for other dimensions of performance.

Conclusion

This report provides a comprehensive performance measurement framework based on publicly reported information on innovative health programs, with operational definitions, best practices, and a user guide. This tool will assist program managers and third-party evaluators in gathering appropriate information for public reports and the CHMI database. It should be reviewed by program managers for usability and investors for relevance. If used to populate the CHMI dataset, it will provide robust information to identify and promote high-performing programs, expanding the reach of successful models of care for the poor.
Introduction

Organizations using market-based approaches to address the health care needs of the poor in sustainable and scalable ways have generated great interest among governments and private investors. The Center for Health Market Innovations (CHMI) aims to collect and analyze data on innovative health programs that seek to provide accessible quality care for the poor in low and middle-income countries. Innovative approaches are inherently risky, as they may not be effective despite their compelling premise or may require significant revision before they can achieve their stated goals. Measuring performance in health services delivery is important for understanding what models work and which should be targeted for expansion or replication. This is of interest to governments, social impact investors, researchers, and program managers. Given the large number of programs that have been identified, common indicators could be developed to highlight high performers. The health services and management literature propose sets of indicators that may be reliable, but may not be feasible for organizations in resource-limited settings to compile. The indicators outlined in this report only describe a part of the work being done by many of the programs, but they include the core elements of organizational performance.

In this report, we review the literature on organizational performance in health care and available information on organizations in the CHMI database to identify best practices in performance reporting. In the previous report, we determined that many of the key fields in the CHMI dataset were not filled, so we supplemented this by collecting information from publicly available sources on 80 of the best described programs. We further refined the performance measurement framework as the data was compiled. We then reviewed the health services and management literature for each framework category, and compared what is recommended in the literature with what is reported by programs. This was used to develop good, better, and best indicators for each field, as well as operational definitions that balance feasibility and credibility, while also considering the comparability of the metric. These definitions have been incorporated into a user guide (see Appendix A), which is intended to assist program managers and third-party evaluators in gathering appropriate information on innovative health programs for the CHMI database. The key deliverables are a set of best practices in performance reporting with a broad hierarchy of validity of performance measures and guidance for future CHMI data collection efforts.

Method

Collecting Supplementary Information from Publicly Available Sources

For this report, we gathered more comprehensive data on a subset of organizations to refine the CHMI framework (which we will call the CHMI Plus framework). In our previous report, we noted the lack of comprehensive data available on most organizations in the CHMI dataset, with information in 41% of all CHMI Plus framework fields. To supplement this data, we focused on extracting data for 80 CHMI programs\(^1\) using the CHMI dataset and profiles, as well as publicly available sources, including program websites and reports, journal articles, and websites located through an online search. We selected programs that had comprehensive information, and programs from one emerging area (mHealth) and one established area (maternal and child health). In total, we extracted comprehensive information on...

\(^1\) In addition, we extracted information on four other programs to collect more data on fields with limited information, specifically utilization, health outcome, cost-to-service ratios, staff-to-service ratios, patient or procedure volume per time period, human resources supply and commitment of staff, political support, and cash flow, ratios, concepts, and calculations.
80 programs, including 35 mobile health (mHealth) programs\(^2\), 49 programs from our top-reporting list\(^3\), 14 well-known programs, and 19 programs focusing on maternal and child health.\(^4\) The process of extracting and classifying the information allowed us to further refine the CHMI framework, adding some categories (such as problem, goal, process, challenges and opportunities, innovations, strategies, and management quality), and breaking down some categories into more specific dimensions (such as availability, quality, efficiency, and sustainability). Quality of amenities was collapsed into other categories since only 6.33% of programs reported this measure. Linguistic/cultural access was also removed since only 6.25% reported this measure. We also used Impact Reporting and Investment Standards (IRIS) as a guide to divide the financial management category into income statement, balance sheet, cash flow, and ratios, concepts, and calculations.

Once we compiled information on 80 programs, we looked across programs to identify good, better, best indicators for each framework category and refined the category definitions to achieve an optimal balance of credibility, feasibility and comparability. For fields containing a large number of possible indicators (specifically, health outcome and impact as well as clinical quality), we have highlighted criteria for assessing indicators rather than developing specific indicators. Categories that require composite indicators or are inferred from multiple sources, such as efficiency and scale, will be discussed in our next report. We also tested the reliability of the definitions and categories through duplicate extraction of data on a subset of programs. The current version of the CHMI Plus framework is described in Table 1, with detailed definitions in Appendix A.

---

\(^2\) According to the WHO (2011, 6) mHealth refers to, “…medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices.” We used this definition to develop a list of 203 mHealth programs in the CHMI dataset, 35 of which were selected as our subset of 80 extracted programs for this deliverable.

\(^3\) The top reporting programs were defined as those with 7 or more of the framework fields filled in original CHMI dataset.

\(^4\) There is considerable overlap of these program areas. For example, 18 mHealth programs are also on our top-reporting list.
### Table 1: CHMI Plus Framework

**Descriptive Fields:**

<table>
<thead>
<tr>
<th>Descriptive</th>
<th>Program Profile</th>
<th>Problem</th>
<th>Goal</th>
<th>Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Challenges / Opportunities</td>
<td>Strategic Planning</td>
<td>Strategies</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**CHMI Innovations**

<table>
<thead>
<tr>
<th>Changing Behaviors</th>
<th>Enhancing Process</th>
<th>Financing Care</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organizing Delivery</td>
<td>Regulating Performance</td>
<td></td>
</tr>
</tbody>
</table>

**Indicator Fields:**

<table>
<thead>
<tr>
<th>Health Status</th>
<th>Population Coverage</th>
<th>Health Output</th>
<th>Health Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Health Access</strong></td>
<td>Affordability</td>
<td>Physical Access and Service Availability / Allocation</td>
<td>Pro-Poor Targeting</td>
</tr>
<tr>
<td><strong>Operations / Delivery</strong></td>
<td>Clinical Quality</td>
<td>User Satisfaction</td>
<td>Management Quality</td>
</tr>
<tr>
<td></td>
<td>Cost-to-Service Ratios</td>
<td>Staff-to-Service Ratios</td>
<td>Patient or Procedure Volume per Time Period</td>
</tr>
<tr>
<td></td>
<td>Commitment of Staff / Human Resources Supply</td>
<td>Operational and Technical Partners</td>
<td>Political Support</td>
</tr>
<tr>
<td></td>
<td>Income Statement</td>
<td>Balance Sheet</td>
<td>Cash Flow</td>
</tr>
<tr>
<td></td>
<td>Ratios, Concepts, and Calculations</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Reporting Practices among Well-Documented Programs

The current CHMI Plus framework contains 19 indicator, 7 descriptive, 5 innovation type, 1 reference, and 1 acronym field. The indicators are grouped by type, with three in the domain category of health status, three for health access, and thirteen for operations/delivery. This framework was used to extract publicly available data on 80 programs listed in the CHMI database, with summary results below.

Aggregating to adjust for changes in the framework, we found 55% overall reporting in indicator fields, compared to 30.3% of reporting in similar fields in the April 30th CHMI Plus framework. Thus, the publicly available information on the best-documented organizations is better than the average for the CHMI dataset, but many key dimensions of performance are still not covered. In Figure 1, the proportion reporting each type of indicator is broken down by field. The top reporting fields are health output (86.25%), and operational and technical partners (82.5%), affordability (71.25%), and balance sheet (68.75%). The poorest reporting fields are staff-to-service ratios (8.75%), cash flows (10%), population coverage (11.25%), and ratios, concepts, and calculations (11.25%).

Figure 1: Percentage Reporting Indicators by Field

When indicators were reported (excluding empty entries), the overall median number of indicators was one. More complex and variable fields had a larger number of reported indicators. Health output, health outcome and impact, and management quality accordingly reported high median number of indicators,

---

5 The operations/delivery indicator fields from the July 15th extractions were aggregated, where appropriate, to allow for comparison on % reporting fields between April 30th and July 15th extractions. Management quality could not be aggregated, but it is averaged into the July 15th % reporting calculation.
with each field reporting a maximum of 37, 24, and 8 indicators in a single program respectively. Figure 2 breaks down the median number reported by field.

Figure 2: Median Number of Reported Indicators by Field

Our framework also includes several descriptive fields. As shown in Figure 3, program profile (100%), problem (76.25%), goal (87.5%), and process (82.5%) were relatively well reported. We were able to collect a median of eight of the possible ten types of data in the program profile.  

Figure 3: % Reporting in CHMI Plus Framework Descriptive Fields

We sought information from publicly available data to supplement what was in the CHMI database to create a comprehensive picture of selected organizations. We found that our expanded approach

---

6 The ten types of data collected in the program profile field are: summary, implementing organization, health focus, type of product/service, country of operation, stage, year launched, number of facilities, number of employees, and target population
yielded more reported data, particularly in the areas of health access and operations/delivery, as summarized in Table 2 below.

Table 2: % Reporting Fields in CHMI Plus framework—CHMI Dataset/Publicly Available Information

<table>
<thead>
<tr>
<th>Field</th>
<th>April 30 CHMI+ using CHMI Dataset</th>
<th>July 15 CHMI+ using CHMI Dataset and Publicly Available Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive/Indicator Fields</td>
<td>41%</td>
<td>50%</td>
</tr>
<tr>
<td>Health Status Indicators</td>
<td>32%</td>
<td>45%</td>
</tr>
<tr>
<td>Health Access Indicators</td>
<td>16%</td>
<td>50%</td>
</tr>
<tr>
<td>Operations/Delivery Indicators</td>
<td>40%</td>
<td>73%</td>
</tr>
</tbody>
</table>

Overview: Evaluating Indicators

Indicators
To assess best practice in performance measurement, we reviewed the literature on indicator development and each dimension in the CHMI Plus framework, and compared it to publicly reported indicators for 80 organizations. Indicators are tools for organizational learning, communication, strategic change, and improvement (Pennsylvania University 2008). Specifically, indicators help in achieving the following (Bastia 2000):

a) Measurement of change: indicators should be sensitive to significant changes in a parameter;
b) Direction: indicators are useful for pointing the direction of change, whether this is positive or negative, whether the situation is improving or worsening; and
c) Measurement of change over time: indicators are designed to measure change over time.

Indicators should be aligned with the objectives and goals of the program as this ensures the indicators are actually measuring what the program has set out to achieve.

Targets/Benchmarks
“Targets are the quantifiable levels of the indicators that a country, society, or organization wants to achieve by a given time” (Kusek and Rist 2004, 91). For example, one target could be “80% vaccination coverage for measles in target population.” Targets for indicators are useful as they indicate the program’s desired level of performance within a specified time period. Benchmarks are standards against which a result is measured. Benchmarking is a systematic process for identifying and implementing best practices (Business Performance Improvement Resource). While setting targets and benchmarks is considered best practice, much of the reporting in the CHMI database did not include targets and/or benchmarks.

Annualizing Data
Programs should report data on an annual basis as this allows one to compare performance of a

---

7 Information sources varied between the April 30th and July 15th extractions, resulting in different sample sizes. Operations/delivery indicator fields were aggregated where appropriate to allow for comparison on % reporting fields between April 30th and July 15th extractions. Management quality could not be aggregated and is omitted from the comparison.

8 For further discussion of good practices for developing robust indicators, see Appendix B, which reviews the standards for CREAM indicators and SMART indicators.
program against a previous year and also facilitates comparability across programs. Data that fails to provide information with an associated time period makes it difficult to track progress and demonstrate the results of a given program. To understand program performance, it is critical to understand whether results have been achieved within one year or over three years. Annualized data allows one to understand how performance varies year over year and whether program results are improving or worsening.

**Ranking Indicators in the CHMI Plus Framework**

In developing a hierarchical ranking of good, better, best indicators for the data in the CHMI Plus framework, our team reviewed the data in each indicator category. Using the relevant literature for each field to determine common approaches to defining the categories, we selected the data that most closely fit these definitions as our good, better, and best indicators. We also evaluated this data using the literature on indicators and their aim to measure change and direction, and facilitate comparison to other programs. The criteria used to rank the indicators as good, better, or best was dictated by the reported data in the field, and different fields employed different criteria to assess indicator ranking. These criteria included ranking based on: the inclusion of a reporting period\(^9\); the inclusion of necessary numerical figures; the level of detail and description (used for qualitative indicators); the inclusion of results from a process; the inclusion of necessary contextual information; the use of an accepted methodology to attain results; the inclusion of a specific target population; and the frequency of accepted standardized processes reported by the program. As well, the reporting of comparative data was also used as a criterion to rank indicators, and this includes comparison of data from multiple reporting periods, inclusion of percentage of change, comparison of program results to that of the wider and/or general population, and comparison of results to those of a similar program. In each category review, we note which criteria shaped the ranking.

For some indicators, multiple time points or controlled designs are needed to determine the impact of a given program, but there was insufficient information in the reported data to assess this. For example, in assessing health outcome and impact, best practices in reporting describe the need for baseline data and the importance of using rigorous designs to assess the changes in health conditions that can be attributed to a particular health intervention. Impact evaluations use methodologies such as experimental, quasi-experimental, and non-experimental designs. Impact assessments can be either qualitative or quantitative and conducted ex ante or ex post intervention, but studies should adopt appropriate, “…timing, sample design and selection, and selection of appropriate survey instruments” (Khandker et al. 2010, 28). Counterfactuals are important to link changes related to the program’s activity. While these components are important to understand health outcomes and impact, and key elements of the quality of the reported data, very few programs provided information on the methodology used to produce their reported results. In addition, impact evaluations require a significant amount of cost, time, and resources to produce, and may not be feasible for many organizations. As a result, other criteria reflecting the reported data were used to rank the indicators in the health outcome and impact category.

\(^9\) IRIS describes the reporting period as the time from the reported Report Start Date to the Report End date (GIIN 2012). We use it here to refer to a defined time period that includes a date.
**Best Practice in Performance Measurement**

**Measuring Population Coverage**

**Literature Review**
In the CHMI framework, utilization is defined as evidence of utilization of a key health intervention at the population level, as reflected by a proportion of a general or target population reached by the program’s health intervention. In the broader health services literature, the term utilization is typically used to report population coverage and facility utilization rate. According to Bradley and colleagues (2010), utilization refers to the volume of services delivered or of clients served, relative to organizational capacity or population health characteristics. However, facility utilization rate in comparison to organizational capacity is a measure of organizational efficiency. The CHMI Plus framework captures facility and organizational capacity and utilization rate under “patient or procedure volume per time period” as a measure of efficiency. Based on the conceptual and empirical literature, we have redefined utilization as population coverage to more precisely articulate what this metric is seeking to capture. According to Murray (2003), population coverage is defined as, “…the proportion of the population in need of an intervention that actually received the intervention” (Murray 2003, 222). This metric is calculated by the number of people who received the intervention divided by the total population who needed the intervention (Murray 2003). The denominator in this calculation is hence the target population for the health intervention.

**Additional Indicator Ranking Criteria**
In addition to the above considerations, the criteria used to select and rank good, better, best indicators for this category were: inclusion of a reporting period, comparison over multiple reporting periods, and inclusion of necessary numerical figures.

**Indicators for Measuring Population Coverage**

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Indicator</th>
<th>Example</th>
<th>Frequency$^{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>Number of clients served as a percentage of a defined target population</td>
<td>Sadguru Netra Chikitsalaya: 80% of tertiary eye care patients in the region are utilizing this service.</td>
<td>7.14%</td>
</tr>
<tr>
<td>Better</td>
<td>Number of clients served as a percentage of a defined target population per reporting period</td>
<td>Top Reseau: In 2002, 7% of youth visited a Top Reseau Clinic.</td>
<td>1.19%</td>
</tr>
<tr>
<td>Best</td>
<td>Comparison of percentages of target population being served showing change over time (including target)</td>
<td>eQuality Health Bwindi: The mountainous rainforest in which the Bwindi hospital is located is home to roughly 60,000 people. When eQuality was first introduced in</td>
<td>1.19%</td>
</tr>
</tbody>
</table>

$^{10}$ The denominator for this reporting frequency is 84 since we engaged in 80 full program extractions to collect data on this field, plus four targeted extractions to collect more information on this and other low-reporting categories.
<table>
<thead>
<tr>
<th>Ranking</th>
<th>Indicator</th>
<th>Example</th>
<th>Frequency¹⁰</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>population number)</td>
<td>February 2010, approximately 2.3% of this population made use of Bwindi’s medical services every month. After the introduction of eQuality, by June 2011, the medical care utilization had doubled to approximately 4.6% of the population per month.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comparison of percentages of target population being served showing change over time (including target population number and number being served)</td>
<td><strong>Hygeia Community Health Plan (HCHP):</strong> The Hygeia Community Health Plan (HCHP) offers health insurance to a total target group of approximately 177,500 people, 67,500 in Lagos and 110,000 in Kwara State. Per July 2010, over 60,000 people were enrolled in the scheme. These figures indicate a 33.8% utilization rate for enrollees per July 2010 data. Holding population levels constant, the March 2011 utilization rate was 42.3% (over 75,000 enrolled).</td>
<td>1.19%</td>
</tr>
</tbody>
</table>

**Total achieving: Good, better, or best** 10.71%

**Recommended Operational Definition**

**Population Coverage:** Volume of clients served as a percentage of a defined target population per reporting period.

**Credibility, Feasibility, and Comparability**

Based on the best reporting standards discussed in the literature, the methods for measuring population coverage are credible. Only 2.38% of the 84 programs reported on population coverage using best practice standards, and this suggests low feasibility. As well, best reporting for this metric is constrained by the fact that programs must possess accurate information on the size of the target population (e.g., pregnant women and children under five). The challenge in precisely defining a target population also limits the comparability of this metric. Reported results that calculate population coverage as defined above can certainly improve comparability; however, the challenge of obtaining population-level data from which target population size can be derived limits the feasibility of this metric. For this category 7.14% of programs reported good data, 1.19% provided better reporting, and 2.38% best reporting.

**Credibility:** 3  
**Feasibility:** 3  
**Comparability:** 2

**Measuring Health Output**

**Literature Review**

According to most of the literature on output measurement, this term refers to quantitative evidence of the health products or services provided as part of a health intervention. The WHO Monitoring and Evaluation framework describes output as pertaining to service provision or delivery, as well as medical...
products and technologies (WHO 2010). Similarly, in evaluating World Bank projects focused on improving reproductive health outcomes, Mills and Bos (2011, 2) define outputs as, “The immediate products or results of the activities implemented, such as the number of personnel trained, number of deliveries conducted, or number of contraceptives distributed.” The CHMI framework follows this approach, defining output as quantitative evidence about the service provided and change in service provision over time (CHMI 2012).

However, the CHMI framework definition for output also includes modeled estimates of impact-based on number of products sold, such as Couple Years Protection (CYPs)\(^\text{11}\) and Disability Adjusted Life Years (DALYs)\(^\text{12}\) (CHMI 2012). This approach differs from the literature on health performance, which considers these outcomes. Indeed, Mills and Bos (2011, 2) describe CYP as an outcome, which is defined as the intermediate result of the implemented activities of a program (Mills and Bos 2011). Whitehead and Ali (2010), Sassi (2006), and Robberstad (2005) also describe DALYs and the related Quality Adjusted Life Years (QALYs)\(^\text{13}\) as measures of health outcome, suggesting that modeled estimates of impact are better captured by the outcome category, an approach that we have adopted.

While the Impact Reporting and Investment Standards (IRIS) framework performance indicators do not include an output category, its product impact category includes many indicators considered to reflect output (for example, indicators describing the organization’s performance and reach) (GIIN 2012). In IRIS’s product impact category, most indicators include the number of patient visits, procedures, or medicinal treatments provided per reporting period. This approach also includes the time dimension of the indicators. In comparison, the Double-Bottom Line framework takes a slightly more open-ended approach, describing outputs as results that a company, nonprofit, or project manager can measure or assess directly (Clark et al. 2004). As well, Kusek and Rist (2004) point out that reporting on measurements collected from different time periods yields important information about the status of a program and can help develop a better understanding of direction and trends over time.

Additional Indicator Ranking Criteria
In addition to the above considerations, the criteria used to select and rank good, better, best indicators for this category were: inclusion of a reporting period, comparison over multiple reporting periods, and inclusion of percentage of change.

### Indicators for Measuring Health Output

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Indicator</th>
<th>Example</th>
<th>Frequency (^\text{14})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>Number of services/products provided or clients</td>
<td><strong>World Health Partners (WHP):</strong> 36,298 tele-consultations conducted between January 2009 and December 31, 2011.</td>
<td>42.5%</td>
</tr>
</tbody>
</table>

\(^\text{11}\) The estimated protection of contraceptive methods during a one-year period calculated based on the volume of contraceptives distributed to clients during that period (Mills and Bos 2011).

\(^\text{12}\) Calculated as years of life lost due to premature mortality in the population and the years of life lost due to disability for incident cases of the health condition (WHO 2012).

\(^\text{13}\) A measure capturing the impact of a treatment on a patient’s health-related quality of life and length of life; it measures the quality of life years gained by an individual from receiving treatments versus no treatment (Whitehead and Ali 2010).

\(^\text{14}\) The denominator for this reporting frequency is 80 since we engaged in 80 full program extractions to collect data on this field.
<table>
<thead>
<tr>
<th>Ranking</th>
<th>Indicator</th>
<th>Example</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>served/trained per reporting period</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Percentage of all specific services/products provided by the particular program per reporting period (market share)</td>
<td>PROSALUD: From 2006-2009, PROSALUD implemented 14% of 116,821 HIV tests implemented nation-wide by the public and private sectors.</td>
<td>1.25%</td>
</tr>
<tr>
<td>Better</td>
<td>Number of services/products provided or clients served/trained per reporting period compared to that of another reporting period</td>
<td>Netra Chikitsalaya: 2002-03 vs. 2010-11: 23,000 surgeries vs. 95,243 surgeries. Health Unit-Concern Universal: PM-Khali clinic children vaccinations in 2007: 2783; Target (3221). PM-Khali clinic children vaccinations in 2008: 1068; Target (1060).</td>
<td>16.25%</td>
</tr>
<tr>
<td>Best</td>
<td>Percentage change in number of services/products provided or clients served/trained per reporting period (including numbers of services/products/ or clients served/trained)</td>
<td>Hygeia Community Health Plan (HCHP): Between December 2007 and December 2010, the schemes recorded a 122% growth in cumulative number of enrollees in the health plan (from 30,019 enrollees in December, 2007 to 66,526 enrollees in December, 2010).</td>
<td>1.25%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total achieving: Good, better, or best</td>
<td></td>
<td>61.25%</td>
<td></td>
</tr>
</tbody>
</table>

**Recommended Operational Definition**

**Health Output:** Quantitative evidence about the number of health services/products provided and/or clients served/trained per reporting period.

**Credibility, Feasibility, and Comparability**

There is a general agreement in the literature on this topic, indicating its strong credibility as a measure. Outputs are credible proxies of outcomes when the procedures described are closely linked with patient or population benefit. As well, it is feasible for programs to track the number of services/products provided and/or clients served/trained during a reporting period, and this should be part of their routine data collection. Also, while this measure provides strong comparability in terms of comparing the number of similar services/products provided or clients served/trained per reporting period, it may not be a useful measure when comparing a wide variety of products and services provided by different programs. For example, the number of children vaccinated may not be a useful comparison to the number of enrollees in a microinsurance health plan. For this category 43.75% of programs reported good data, 16.25% provided better reporting, and 1.25% best reporting.
Measuring Health Outcome and Impact

Literature Review
The CHMI framework defines health outcome as, “...evidence of impact as demonstrated by improvements in health indicators” (CHMI 2012). The literature distinguishes between health outcome and impact. Health outcomes are intermediate observable and measurable changes, which include, “...specific changes in attitudes, behaviors, knowledge, skills, status, or level of functioning that result from enterprise activities” (Clark et al. 2003; Twersky et al. 2010). Outcome indicators depend on the intended goals/objectives of the intervention (e.g. condom usage). Outcome measures include components of learning as reflected by changes in awareness, knowledge, attitudes, skill development, and motivation, as well as actionable components such as changes in behavior, practices, decision-making, and actions (Taylor-Powell et al. 2003).

Health impact refers to long-term health outcomes as well as changes in the social, economic, civic and environmental conditions as a consequence of the program (Mills and Bos 2011; Taylor-Powell et al., 2003). Some examples of health impact include: change in mortality rates and calculated measures of health outcome such as disease-adjusted life years (DALYs) and quality-adjusted life years (QALYs) (Jee and Or 1999). Kelley and Hurst (2006) distinguish between outcomes and impacts where, “...intermediary outcome indicators assess the short term outcome while the final outcome indicators assess the long term outcome.” Some literature on outcome includes patient satisfaction, but the CHMI Plus framework captures this in the user satisfaction category.

Additional Indicator Ranking Criteria
In addition to the above considerations, the criteria used to select and rank good, better, best indicators for this category were: inclusion of a reporting period, comparison to a similar and/or wider population base, and comparison over multiple reporting periods.
## Criteria for Measuring Health Outcome and Impact

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Criteria</th>
<th>Example</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>Outcome measure with respect to the intervention per client base per reporting period</td>
<td><strong>PROSALUD</strong>&lt;br&gt;In 2009, PROSALUD centers achieved an average 93% tuberculosis cure rate. This exceeds the WHO global target cure rate of 85%.&lt;br&gt;&lt;br&gt;<strong>Sun Quality Health Network, Myanmar</strong>&lt;br&gt;Franchising will produce 90,000 of the program’s 130,000 total DALYs in 2010 at an average cost per DALY of around USD $70. TB treatment provides the majority of those DALYs at a cost per DALY of approximately USD $40.&lt;br&gt;-CYPs: 2009 - 255754; 2010 - 263740; 2011 - 280500&lt;br&gt;-DALYs averted: 2010: SRH: 19,040; HIV/AIDS: 2,796&lt;br&gt;Mean Corpuscular Hemoglobin: 6,801; Malaria: 18,523; TB: 57,986</td>
<td>33.75%</td>
</tr>
<tr>
<td>Better</td>
<td>Outcome measure per target population or per client base compared to similar and/or wider population base per reporting period</td>
<td><strong>eQuality Health Bwindi</strong>&lt;br&gt;The maternal mortality rate in Uganda is 0.44% (UNICEF, 2009). The rate in the maternity ward of Bwindi’s Hospital is more than four times lower in FY 2010-2011 at 0.10%.&lt;br&gt;TB treatment default rate ranges from 11.3% to 29.6% across sub-Saharan Africa (B.Castelnuovo, 2010). Bwindi's default rate is 3.3%.</td>
<td>7.50%</td>
</tr>
</tbody>
</table>

---

15 The denominator for this reporting frequency is 84 since we engaged in 80 full program extractions to collect data on this field, in addition to four targeted extractions to collect more information on this and other low-reporting categories.
<table>
<thead>
<tr>
<th>Ranking</th>
<th>Criteria</th>
<th>Example</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best</td>
<td>Outcome measure per target population or per client base compared to target population over multiple reporting periods</td>
<td><strong>Karuna Trust</strong>&lt;br&gt;In states served by Karuna Trust Primary Health Care System from 1996 to 2007:&lt;br&gt;<em>Infant Mortality Rate dropped from 75.7% to 23.8%</em>&lt;br&gt;<em>Still birth rate: 37.8% to 10.2%</em>&lt;br&gt;<em>Perinatal Mortality Rate: 67.7% to 17%</em>&lt;br&gt;<em>Neonatal Mortality Rate: 70.3% to 10.2%</em>&lt;br&gt;<em>Child Mortality Rate (1-5 years): 12.4% to 3%</em>&lt;br&gt;<em>Under-5 Mortality Rate: 88.1% to 26.8%</em>&lt;br&gt;<em>&gt;85% of TB patients cured with treatment</em>&lt;br&gt;<em>100% Immunization coverage for pregnant women in 2007-2008</em>&lt;br&gt;<em>97% Children immunization</em>&lt;br&gt;<em>95.4% Institutional delivery rate</em></td>
<td>5.00%</td>
</tr>
</tbody>
</table>

**Total achieving: Good, better, or best** 46.25%

**Recommended Operational Definition**

**Health outcome and impact**: Quantitative evidence of intermediate or long-term health outcome evidenced by changes in learning, actions, or health status of clients/target population per reporting period.

**Credibility, Feasibility, and Comparability**

Literature on health outcome and impact agrees that health outcomes are intermediate results while health impacts are long-term results. Many sources in the literature lack more precise definitions of each, instead providing examples in its place. Good practice in health outcomes reporting per client base requires continuous monitoring. In terms of feasibility, better and best practices in health outcomes require collecting information on target population and baseline data, which will pose a challenge to programs with limited resources. Certain indicators should be risk-adjusted (e.g. mortality rate) to assess a program on controllable factors. Risk adjustment may be beyond the capacity of resource-constrained programs, but without it, the credibility of the indicator is undermined. Impact assessments have similar challenges, demanding more significant cost, time, and resource commitments. The examples of outcome measures given above, especially standard measurements of health outcome such as mortality rates, CYPs, DALYs and QALYs, are highly comparable. Investors can use these standard measures to assess the scope and effectiveness of health service organizations within the same service field (i.e. family planning organizations can be compared using CYPs). Impact evaluations should be tailored to specific interventions, which may limit comparability across programs. For this category 33.75% of programs reported good data, 7.50% provided better reporting, and 5.00% best reporting.
Measuring Affordability

Literature Review
Affordability is a measurement of the ability of particular consumer groups to pay for a minimum level of a certain service. The affordability category was originally termed cost (in the original CHMI framework), which Bradley and colleagues (2010) consider a measure of accessibility. Reducing cost is a core feature of improving affordability, and is integral to high-volume, low-cost business models and cross-subsidy strategies employed by health service delivery organizations in low and middle-income countries (Bhattacharyya et al. 2011). The CHMI Reported Results framework uses the following approach for measuring affordability: “Evidence of a change in the price of products or services charged to the patient. Statements given must either show a change in cost over time or present evidence of higher/lower costs compared to the price of equally valued good/services elsewhere” (CHMI 2012). In the CHMI Plus framework, the cost field has been renamed affordability in an effort to better capture this aim of increasing financial access to health products and services.

Affordability is also a measure of health care access, and is a person’s ability to access care without the risk of becoming impoverished by its cost. While there are many programs that support subsidizing the delivery of care, insurance programs work to prevent catastrophic costs in the event of an acute health event. Dror and Jacquier (2001) refer to affordability as indicators that describe the ability of a client to pay the insurance premium, the periodicity of premium payments, comprehensiveness of the package and finally the type of transaction that occurs when the insured accesses care (deductibles, co-payments, reimbursement of costs, or direct coverage).

Additional Indicator Ranking Criteria
In addition to the above considerations, the criteria used to select and rank good, better, best indicators for this category were: comparison over multiple reporting periods, comparison of program offerings to similar offerings, and the inclusion of necessary contextual criteria/information.

Indicators for Measuring Affordability

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Indicator</th>
<th>Example</th>
<th>Frequency 16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>Price of service and/or product per unit</td>
<td>HealthLine Bangladesh: The call charges are fixed at BDT 15.00 for 3 minutes (about US $0.21).</td>
<td>30.00%</td>
</tr>
<tr>
<td></td>
<td>Free</td>
<td>GlicOnline: As conceptualized, GlicOnLine is free for doctors and patients (except for conventional telephone costs to access the application).</td>
<td>28.75%</td>
</tr>
<tr>
<td></td>
<td>Savings accrued from voucher or subsidizing scheme</td>
<td>Top Réseau: Voucher decreases cost of accessing family planning services by half</td>
<td>15.00%</td>
</tr>
</tbody>
</table>

16 The denominator for this reporting frequency is 80 since we engaged in 80 full program extractions to collect data on this field.
<table>
<thead>
<tr>
<th>Ranking</th>
<th>Indicator</th>
<th>Example</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Price of service and/or product compared to similar non-local product and/or service; or compared to related local service and/or product it replaces</td>
<td><strong>Teledoctor:</strong> Using the TeleDoctor service, patients can get a tele-consultation with a doctor for call charges of only 10 Indian rupees (INR) for three minutes, compared to the cumulative cost of 200 INR including a minimum consultation fee of 100 INR and average transportation cost of 100 INR to access traditional services.</td>
<td>5.00%</td>
</tr>
<tr>
<td></td>
<td>Price of service compared to perceived value</td>
<td><strong>Health Management and Research Institute:</strong> The value of the Health Information Helpline triaging services, received free of charge, is perceived by the callers at more than 3 billion INR.</td>
<td>1.25%</td>
</tr>
<tr>
<td>Better</td>
<td>Price of service and/or product, compared with costs of accessing the same service or product locally</td>
<td><strong>PROSALUD:</strong> Charges approximately USD 4.00 for an appointment with a general practitioner, compared to USD 28.00 in the private sector. <strong>SalaUno Salud:</strong> Prices are 1/3 the market price of the closest competitor due to the innovative Small Incision Cataract Surgeries (SICS) and the ability to take advantage of economies of scale utilized in a high volume treatment model.</td>
<td>21.75%</td>
</tr>
<tr>
<td></td>
<td>Change in price over time to improve affordability.</td>
<td><strong>Rashtriya Swasthya Bima Yojana:</strong> The average premium has decreased over time from 600 INR per family per year in 2008 to 500 INR in 2009.</td>
<td>5.00%</td>
</tr>
<tr>
<td></td>
<td>Price of service and/or product expressed as a proportion of income</td>
<td><strong>Pesinet:</strong> Families subscribe for this service on a voluntary basis. The monthly price for the whole package of services is 500 FCFA (USD 1) per child, the equivalent of a kilo of onions, a price affordable to low-income families in Bamako. Subscribers pay an affordable monthly fee for the service – less than 1€, the equivalent of a day of work for 90% of Malian people.</td>
<td>1.25%</td>
</tr>
<tr>
<td>Best</td>
<td>Price of service and/or product related to ability to pay of the poorest group</td>
<td><strong>Sun Quality Health Network, Myanmar:</strong> Population Service International (PSI) Myanmar ascertained that doctors considered that they needed to receive approximately $5 USD to perform an intrauterine device insertion in order for the service to be a viable part of their practice. PSI Myanmar also determined that the patient population could only afford to pay $0.50 USD for this procedure. To address</td>
<td>2.50%</td>
</tr>
</tbody>
</table>
these issues PSI Myanmar started a program to reimburse the providers the remaining $4.50 USD for each IUD inserted.

**Total achieving: Good, better, or best**

110.50%

**Recommended Operational Definition:**

**Affordability:** Quantitative evidence about the price of services and/or product in comparison to the average cost of accessing similar services in the local context or as a proportion of income.

**Credibility, Feasibility, and Comparability:**

For this category 80.00% of programs reported good data, 26.75% provided better reporting, and 3.75% best reporting. From the current database, the credibility of indicators is based on relating the cost of the product or service to the income of the recipient group. However, from the database, this seems to be an infeasible measure to collect, or report. Furthermore, it becomes difficult to compare, when the comparison tool used is variable (i.e. a kilo of onions, vs. the ability to pay). Taking this into account, the most feasible indicators include reporting on the current price that programs charge their clients (29% of programs report this), and including the comparison of the market prices in the local context (19% of programs report this). Comparability of these indicators is difficult with the best level indicators; however it can be accomplished with the better and good categories. As a result, our recommended definition, which involves comparison of price of product and/or service to local market prices, is considered somewhat comparable across different program offerings.

![Credibility: ★★★ Feasibility: ★★★ Comparability: ★★★](image)

**Measuring Physical Access and Service Availability / Allocation**

**Literature Review**

The CHMI framework defines availability as, “...evidence of a change in the availability of services in previously underserved areas” (CHMI 2012). The Bradley framework (2010) groups indicators on access into six categories: physical, linguistic, information, service availability/allocation, non-discriminatory service provision, and affordability. Similarly, the WHO Monitoring and Evaluation framework measures access based upon physical, economic, and socio-psychological aspects of people’s ability to make use of health services (WHO 2010). The CHMI Plus framework addresses dimensions of financial access under affordability, and consequently we have excluded this from the access category. Kelley and Hurst (2006,

---

17 The “total achieving: good, better, best” can include multiple indicators per program, resulting in a figure greater than 100%.
26) emphasize the time aspect of access, stressing that clients/patients should be able, “...to obtain care/service at the right place and right time.”

**Additional Indicator Ranking Criteria**

In addition to the above considerations, the criteria used to select and rank good, better, best indicators for this category were: inclusion of reporting period, comparison over multiple reporting periods, inclusion of a target population, and comparison of program offerings to similar offerings.

**Indicators for Measuring Physical Access and Service Availability / Allocation**

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Indicator</th>
<th>Example</th>
<th>Frequency$^{18}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>Number of facilities, providers, or hospital beds per population per reporting period</td>
<td>SMS for Life: Public-private initiative ran from September 2009 to February 2010. Three districts in Tanzania - Lindi Rural, Kigoma Rural and Ulanga - were involved, covering 129 health facilities and 226 villages, representing 1.2 million people (i.e. 1 facility per 9302 people).</td>
<td>5.00%</td>
</tr>
<tr>
<td></td>
<td>Geographic distance to facility or provider per reporting period</td>
<td>Operation ASHA: Each one of our urban treatment centers serves anywhere from 5,000-25,000 patients within a 1.5 km radius of the center.</td>
<td>5.00%</td>
</tr>
<tr>
<td></td>
<td>Number and % absenteeism per reporting period</td>
<td>Initiative on Primary Healthcare (IPH): Significantly reducing absenteeism by as much as 96% in 2009</td>
<td>1.25%</td>
</tr>
<tr>
<td></td>
<td>Number and % of stock-outs per reporting period</td>
<td>SMS for Life: The proportion of health facilities with no stock of one or more antimalarial medicine fell from 78% at week 1 to 26% at week 21</td>
<td>1.25%</td>
</tr>
<tr>
<td></td>
<td>Facility or provider hours of service per reporting period</td>
<td>Health Unit-Concern Universal: Each of the Kolyani clinics is open from 9:00 a.m. to 5:00 p.m., six days a week</td>
<td>5.00%</td>
</tr>
<tr>
<td>Better</td>
<td>Number of facilities, providers, or hospital beds per target population per reporting period</td>
<td>N/A</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>Geographic distance and time taken to travel to facility or provider by mode of transportation per reporting period</td>
<td>Operation ASHA: Each one of our urban treatment centers within a 1.5 km radius of the center. Our centers are placed so that no patient has to walk more than 10 minutes to reach the center.</td>
<td>1.25%</td>
</tr>
</tbody>
</table>

$^{18}$ The denominator for this reporting frequency is 80 since we engaged in 80 full program extractions to collect data on this field.
<table>
<thead>
<tr>
<th>Ranking</th>
<th>Indicator</th>
<th>Example</th>
<th>Frequency&lt;sup&gt;18&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Change over time in number and % absenteeism</td>
<td>N/A</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>Change over time in number and % of stock-outs</td>
<td>N/A</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>Comparison of facility or provider hours of service with similar facility or provider per reporting period</td>
<td>Operation ASHA: Most other Directly Observed Treatment Short-Course (DOTS) treatment centers mostly stay open an average of 4 to 6 hours between 10 a.m. and 4 p.m. during the day. OpASHA centers open early in the morning until late in the evening</td>
<td>1.25%</td>
</tr>
<tr>
<td>Best</td>
<td>Change over time in number of facilities, providers, or hospital beds per target population</td>
<td>N/A</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td><strong>Total achieving: Good, better, or best</strong></td>
<td></td>
<td><strong>20%</strong></td>
</tr>
</tbody>
</table>

**Recommended Operational Definition**

**Physical Access and Service Availability:** Quantitative evidence about the ability of clients/patients to access health services at the needed place and time per reporting period. Measures include: number of facilities, providers or hospital beds per segment of population or target population, distance to facility, staff absenteeism, medical supply stock-outs, and hours of operation.

**Credibility, Feasibility, and Comparability**

The sample extraction has 17.50% good, 2.50% better, and 0% best indicators. Physical access and service availability have moderately comparable data, but the utility of comparison is limited to programs addressing a similar clinical problem. While data should include reporting period, which provides context to the data, our sample did not exclude indicators on this basis to allow for evaluation of type of indicators reported. While target population data may be difficult to regularly collect, gathering data on the other measures of physical access and service availability is more feasible.

![Credibility: 3/3](image)

![Feasibility: 3/3](image)

![Comparability: 3/3](image)

**Measuring Pro-Poor Targeting**

One of the core challenges of health services program is effectively reaching the poor. Understanding how programs deliver to the poor is important as the lack of clear targeting criteria can lead to leakage (non-poor being included inadvertently) or poor coverage (many poor people not being identified).
**Literature review**

The frameworks measuring organizational performance in health services delivery that were previously examined do not provide information on pro-poor targeting. Kelley and Hurst (2006) include equity as a commonly used dimension in quality of care frameworks. Bradley and associates (2010) identify access as a core performance domain and within this financial access is a component. These frameworks however do not specify performance criteria for pro-poor targeting in health services delivery.

There are a plethora of analytical frameworks for understanding and measuring poverty. Poverty can be measured using service consumption or income or access to services at the households and/or community levels. One multidimensional method of measuring of poverty is the asset index approach. Rather than comparing expenditures, asset indices capture the evolution of assets over time. Conceptually, assets reflecting productive capacity form a more robust basis for identifying the poor than do flow variables such as expenditures or income (Liverpool and Winter-Nelson 2010). Asset indices offer a rigorous method for pro-poor targeting due to their high quality and standardized data collection scheme. Proxy measures of poverty such as belonging to a disadvantaged group (e.g. Scheduled Castes) or by type of job (e.g. informal laborers) are also useful ways to target the poor. While asset indices offer a credible means for pro-poor targeting they are not particularly feasible. Proxy measures, which rank lower in terms of credibly, are more feasible to employ in pro-poor targeting.

An important issue is whether the target group should be defined in absolute or relative terms (Duclos 2009). Absolute poverty refers to the category of people who are unable to meet the basic threshold that is needed for survival. Relative poverty occurs when people are poor when compared to others around them (Hulme, Moore and Shepherd 2001). The choice of which measure to use is highly dependent on the context. For example, in some countries, it may be appropriate to use the minimum wage or the minimum requirements for physical well-being based on a quantitative proxy indicator such as income or calories (Hulme, Moore and Shepherd 2001).

Patouillard, Goodman, Hanson and Mills (2007) effectively operationalize the concepts of absolute and relative poverty in the context of health services delivery. The authors assert that if an intervention is being implemented in a poor area, then the intervention is pro-poor; this approach is premised on the concept of absolute poverty. The second approach is aligned with the concept of relative poverty and maintains that if the socio-economic distribution of benefits favors the most disadvantaged group within a population, then the intervention is pro-poor. Both of these approaches demand a consideration of the geographic context in which the health intervention is being implemented. The Niger delta region for example, is marked by widespread deprivation and underdevelopment. In this context, using an absolute measure of poverty is more relevant as the majority of the population is poor. Alternatively, research shows that there is a strong overlap between being a women informal worker and being poor. Thus, health interventions targeting women informal workers, a socio-economically disadvantaged group compared to the rest of the population would be considered a relative pro-poor targeting strategy.

**Additional Indicator Ranking Criteria**

In addition to the above considerations, the criteria used to select and rank good, better, best indicators for this category were: inclusion of a reporting period and the inclusion of necessary contextual criteria/information.
## Indicators for Measuring Pro-Poor Targeting

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Indicator</th>
<th>Example</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>Assessment criteria</td>
<td><strong>Bangladesh Demand Side Financing:</strong> The following criteria were used to identify the target women: resident of the sub-district; current pregnancy is the first or the second pregnancy; functionally landless; i.e. owing less than 6534 square feet of land, household earning less than US $38.50 per month and lack of ownership of other productive assets.</td>
<td>3.75%</td>
</tr>
<tr>
<td></td>
<td>Income level per time period</td>
<td><strong>Greenstar:</strong> Household poverty status was determined by rank of monthly income, with those households in the twentieth percentile or below classified as poor.</td>
<td>2.50%</td>
</tr>
<tr>
<td></td>
<td>Proportion of clients</td>
<td><strong>Lifespring:</strong> The approximate monthly family income of customers is as follows: 40% earn below $90 per month; 30% earn between $91 and $130 per month; 20% earn between $131 and $220 per month; and 10% earn above $220 per month.</td>
<td>2.50%</td>
</tr>
<tr>
<td></td>
<td>Proportion of clients per reporting period</td>
<td><strong>Health Management and Research Institute (HMRI):</strong> All social groups are utilizing the service, though 38% of the services are utilized by socially deprived populations like Scheduled Castes and Scheduled Tribes.</td>
<td></td>
</tr>
<tr>
<td>Better</td>
<td>% clients/patients below poverty line</td>
<td><strong>CARE Hospitals:</strong> In 2009-10, the overall patient mix consisted of 25 percent in the low income group, 50 percent in the middle income group and 25 percent in the upper income group.</td>
<td>1.25%</td>
</tr>
<tr>
<td></td>
<td>Benefit incidence</td>
<td><strong>Bangladesh Demand Side Financing:</strong> Per 2011 data, rich to poor ratio for % of women receiving antenatal care at least 3 times from trained providers: 1.5; rich to poor ratio of % of deliveries assisted by skilled health personnel: 1.3; rich to poor ratio of % babies delivered at a health facility: 1.6; ratio of rich to poor % receiving postnatal care from a trained providers within 42 days of delivery.</td>
<td>1.25%</td>
</tr>
</tbody>
</table>

The denominator for this reporting frequency is 80 since we engaged in 80 full program extractions to collect data on this field.
### Pro-poor Targeting: Universal Health Unit

In the reporting period Jan-Dec 2008, the project classified targeted families by their individual economic conditions using the following criteria:

1. **Solvent:** those with:
   - a) household monthly earnings in excess of 2,500 Bangladesh Taka (BDT) (equivalent to 18 pounds);
   - b) land for cultivation;
   - c) paying land and income taxes;

2. **Poor:** those with household monthly income less than 2,500 BDT;

3. **Poorest (Hard core poor):** those that meet at least two of the following:
   - a) household income of under 1,500 BDT,
   - b) without resources for income generation,
   - c) solely dependent on jhum (shifting slash and burn) cultivation;

   Additionally, anyone living in a hard to reach area was automatically classified within the 'poorest' category.

*As per CORAID report covering reporting period Jan-Dec 2008:

Target population came from 28,544 households of whom 4,202 were classified as solvent, 11,728 as poor, and 12,545 as hard core poor.

### Total achieving: Good, better, or best

<table>
<thead>
<tr>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.25%</td>
</tr>
</tbody>
</table>

---

**Recommended Operational Definition**

**Pro-poor Targeting:** The extent to which the program’s clients are disadvantaged in terms of socio-economic status and the criteria used to determine and target the poor. This includes reference to whether the targeting is being done in a predominately poor area or is targeting the most disadvantaged group within a population.

**Credibility, Feasibility and Comparability**

The choice of which measure to use (i.e. absolute or relative) is dependent on the context. Targeting the poor is easier if everyone in a region is poor. Since twice as many of the world's poor live in 'middle-income' countries as in 'low-income' countries, targeting the poorest is more important in settings where average incomes are higher (University of Oxford, n.d.). The recommended operational definition for this category offers a feasible way to report for this category. The comparability of pro-poor targeting across organizations is challenging as the criteria and methods used may vary considerably. For this category 8.75% of programs reported good data, 6.25% provided better reporting, and 1.25% best reporting.
Measuring Clinical Quality

Literature Review
The Institute of Medicine (2001) defines clinical quality as, “...the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge.” Accordingly, Bradley et al. (2010, 3), refer to clinical quality as, “...whether the provider’s care conformed to best clinical practice for those who use the services of the organization; it does not refer to outcome measures of population health, such as vaccination or antenatal care coverage, in which the denominator is the population.” Clinical quality indicators thus include the ability of providers to deliver best-practice care, which is context-sensitive to the resources and the setting where they practice. For example, the International Diabetes Federation’s guidelines for diabetes in childhood and adolescence vary based on a region’s ability to provide standards of care with limited resources (IDF 2011).

The National Health System (NHS) of the United Kingdom’s long history of quality improvement, and Darzi’s (2008) High Quality Care for All, set out three main quality dimensions: effectiveness, patient experience, and safety. The NHS has thus developed expansive lists of key clinical indicators that are specific to practice areas. Instead of replicating the list of indicators, the CHMI Plus framework identifies components that aid in ranking the quality of indicators according to the central domains including effectiveness of care, safety of care, and meeting appropriate standards of care.

The CHMI Plus framework encompasses the governing concepts of the literature definitions and their components through several dimensions. The particular dimensions of clinical quality in this framework therefore refer specifically to indicators that measure quality of health service provision for individuals and populations that results in safe, effective, appropriate evidence-based care supported with comparators and a time frame.

Additional Indicator Ranking Criteria
In addition to the above considerations, the criteria used to select and rank good, better, best indicators for this category were: inclusion of a reporting period, comparison over multiple reporting periods, and comparison of program offerings to similar offerings.

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Criteria</th>
<th>Example</th>
<th>Frequency&lt;sup&gt;20&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>Quantitative evidence of adherence to practices and procedures based on clinical guidelines/standard of care</td>
<td>ACQUIRE Tanzania Project: 90% of women getting prenatal care also received voluntary HIV testing.</td>
<td>5.00%</td>
</tr>
</tbody>
</table>

<sup>20</sup> The denominator for this reporting frequency is 80 since we engaged in 80 full program extractions to collect data on this field.
<table>
<thead>
<tr>
<th>Ranking</th>
<th>Criteria</th>
<th>Example</th>
<th>Frequency²⁰</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Quantitative evidence of the ability of the intervention to provide</td>
<td>N/A</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>effectiveness of care</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Better</td>
<td>Quantitative evidence of the safety of the intervention (includes error</td>
<td>Enhanced Diarrheal Disease Control in Vietnam: 2.1% of children receiving treatment experienced mild adverse responses.</td>
<td>1.25%</td>
</tr>
<tr>
<td></td>
<td>and complication rates)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quantitative evidence of the effectiveness of care per reporting period</td>
<td>Operation ASHA: The tuberculosis detection rate in areas served by Operation ASHA has increased by 95% over the four years that Operation ASHA has been working in these areas.</td>
<td>1.25%</td>
</tr>
<tr>
<td></td>
<td>Comparative quantitative evidence of effective care</td>
<td>Sadguru Netra Chikitsalaya: In 2010, 98% of 35776 cataract surgeries include an intraocular lens vs. 55% of 23525 in 2000</td>
<td>3.75%</td>
</tr>
<tr>
<td></td>
<td>Comparative quantitative evidence of the safety of the intervention</td>
<td>Aravind Eye Care System: Despite the shared spaces, Aravind has managed to keep its infection rates low, an average of about 4 cases per 10,000 patients, compared to an average 6 per 10,000 in the U.K.</td>
<td>2.50%</td>
</tr>
<tr>
<td>Best</td>
<td>Comparative quantitative evidence of adherence to guidelines/standard of</td>
<td>Centre Dushisho: Comparing indicators collected at the 4 Dushisho Centers for a 9 month period before and after Family planning (FP)/HIV integration was strengthened (in July 2009), we observed an increase in the proportion of Voluntary counseling and testing (VCT) clients who received FP information (40% versus 93%) and proportion of VCT clients who were referred for FP services at a health facility and/or accepted condoms (69% versus 90%).</td>
<td>1.25%</td>
</tr>
<tr>
<td></td>
<td>care per reporting period</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total achieving: Good, better, or best 18.75%

Recommendations/Operational Definition:

Clinical Quality: Quantitative evidence of providing safe, effective, evidence-based care, compared to other providers of similar services, or demonstrates change over time.
Credibility, Feasibility, and Comparability:
The credibility for clinical quality indicators is strengthened when the indicator measures a change in clinical quality over time or provides clinical quality performance as compared to another program. Performance compared to other organizations or compared to another reporting period can display the robustness of a systematic clinical process that yields quality, or conversely, identify if a result was due to chance. According to the framework, clinical quality is not a well-reported statistic where only 13 out of 80 programs reported any clinical quality measure. This may speak to the challenges program managers face in measuring clinical quality given the resources required for this assessment, which can include access to current knowledge of best clinical practices and sophisticated measurement and calculation of adherence to these standards. Finally, this measure can serve as a useful comparator, between programs that deliver similar services. For this category, 6.25% of programs reported good data, 5.00% provided better reporting, and 7.50% provided best reporting.

Credibility: ★★★ Feasibility: ★★★ Comparability: ★★★

Measuring User Satisfaction

Literature Review
According to the CHMI Reported Results framework, user satisfaction is, “…evidence of good service quality as perceived by the patient” (CHMI 2012). User satisfaction is generally considered an outcome quality measure in the literature (Darzi 2008; National Council on Quality Assurance 2012). Donabedian (1988) identifies the quality of interpersonal care as a key aspect of patient satisfaction. Kelley and Hurst (2006) explore the concept of patient-centered care and the responsiveness of healthcare staff as a measure of patient satisfaction. Bradley et al. (2010) also consider patient satisfaction as a measure of quality, although they recognize the highly subjective nature of this field given that it is related to individual expectation. Kaplan and Norton (1992) define user satisfaction as a patient’s perception of the quality of services provided by reporting on their experiences, the evaluation of the services they received, and their interaction with hospital staff. Some examples include overall impressions, communication, access to care and services, consideration and respect, continuity and transition, coordination and integration, information, education and communication. Crow et al. (2002) refer to satisfaction as user-perceived quality of services or user experiences and identify patient-practitioner relationship as the most important health service factor affecting satisfaction. A related aspect of user satisfaction is patient experience, which the Beryl Institute (2012) defines as, “...the sum of all interactions, shaped by an organization’s culture, that influence patient perceptions across the continuum of care.” Patient experience is a crucial component of user satisfaction; however, based on the type of data reported by CHMI programs, we have focused on user satisfaction, which includes an aspect of patient expectation and evaluation in the patient experience.

The literature identifies two types of questions to elicit user satisfaction: direct and indirect. Direct questions explicitly ask about levels of satisfaction with care experience (Aspinal et al. 2003). For example, this includes the following question: On a scale of 1-5, how would you rate your satisfaction with the care you received from the nurse, 5 being very satisfied, and 1 being disappointed (Aspinal et al. 2003). Indirect questions are measures that expect the researcher to infer the user satisfaction from questions that assess usage, or repeat usage of services. The IRIS framework utilizes this type of measure when it includes client retention rate as a metric of product impact (GIIN 2012). The IRIS
framework also highlights user satisfaction activities as an indicator for information and learning (GIIN 2012). Finally, a survey completed by 2160 adults found that higher patient satisfaction is significantly associated with patient-initiated referral of physicians, and a simple question regarding recommendation of the doctor can be used as a surrogate measure of patient satisfaction, although it should be interpreted with caution (Kersnik 2003, 247-250).

There are several validated tools that measure patient satisfaction such as the Primary Care Assessment Survey (PCAS; Safran et al. 1997), the Primary Care Assessment Tools (PCAT; Berra et al. 2011), and the General Practice Assessment Survey (GPAS; Ramsay et al. 2000). Tools that assess patient experience and satisfaction can be modified to suit the needs of a particular practice and patient population. The 80 extracted programs did not measure satisfaction using validated tools supported by literature; however, some developed and utilized their own systematic assessment method such as structured interviews and surveys to determine user satisfaction.

**Additional Indicator Ranking Criteria**
In addition to the above considerations, the criteria used to select and rank good, better, best indicators for this category were: the use of an accepted methodology to attain results.

### Indicators for Measuring User Satisfaction

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Indicator</th>
<th>Example</th>
<th>Frequency&lt;sup&gt;21&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Good</strong></td>
<td>User satisfaction quantitatively reported, without inclusion of a systematic assessment method</td>
<td><strong>Learning about Living</strong>: 76% of SMS users were satisfied with the service, and over half of the &quot;unsatisfied&quot; users cited problems likely due to the congested telecom networks (rather than the OneWorld service) as the reason for their disappointment.</td>
<td>5.00%</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Pesinet</strong>: 99% want to recommend the service to their relatives or friends.</td>
<td>1.25%</td>
</tr>
<tr>
<td><strong>Better</strong></td>
<td>User satisfaction quantitatively reported, with a systematic assessment method</td>
<td><strong>Mahila Swastha Sewa</strong>: Exit interviews conducted in 2010 show that 80% of clients reported being satisfied with the services provided.</td>
<td>8.75%</td>
</tr>
<tr>
<td></td>
<td>Retention Rates or re-subscription rates.</td>
<td><strong>Simling Sun Franchise Program</strong>: The clients are typically loyal users of SSFP services and the franchise found that that 71 percent of SSFP customers are repeat users.</td>
<td>6.25%</td>
</tr>
</tbody>
</table>

<sup>21</sup> The denominator for this reporting frequency is 80 since we engaged in 80 full program extractions to collect data on this field.
<table>
<thead>
<tr>
<th>Ranking</th>
<th>Indicator</th>
<th>Example</th>
<th>Frequency $^{21}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best</td>
<td>User Satisfaction quantitatively reported with a systematic methodology by a third party</td>
<td><em>Pesinet</em>: An evaluation survey carried out by a PhD student under the supervision of the National Agency for e-health and Medical Informatics in Mali shows that 96% of the families are satisfied with the service.</td>
<td>6.25%</td>
</tr>
</tbody>
</table>

**Total achieving: Good, better, or best** 27.5%

**Recommended Operational Definition**

**User Satisfaction**: Quantitative or qualitative evidence that is collected using a systematic methodology and reflects the client’s perception on the quality of services provided.

**Credibility, Feasibility, and Comparability**

According to the available literature, there is general agreement on the definition of user satisfaction as a patient-perceived quality measure, which adds credibility to our operational definition. User satisfaction was reported by 27.5% of the programs in the database, suggesting that the measurement and reporting of user satisfaction is reasonably feasible. Systematically measuring and reporting the data is also relatively feasible since 8.75% were able to do this at a better indicator level and 6.25% reported on this at a best level, which included third-party evaluation. However, some programs may lack the support necessary to develop strong methodology in measuring and monitoring user satisfaction and may not have the resources for third party evaluation, which is why we simply encourage the use of a systematic methodology in our recommended definition. In addition, most programs develop their own systematic methodology in measuring user satisfaction, which can lower comparability. For this category 6.25% of programs reported good data, 15.00% provided better reporting, and 6.25% best reporting.

**Credibility:** 🍃🍃🍃 **Feasibility:** 🍃🍃🍃 **Comparability:** 🍃🍃🍃

**Measuring Management Quality**

In a context where organizations are being increasingly pressured to simultaneously plan and allocate resources to ensure high quality standards while reducing costs, management quality becomes a critical component warranting attention in the health services industry. The normative and operational frameworks of health systems recognize that management quality plays a critical role in health system performance.

**Literature review**

The literature on Total Quality Management (TQM) documents that quality management is a vital ingredient to strategic planning within the health care domain (Anderson and Zwelling, 1996). In *Developing Strategies for Improving Health Care Delivery*, managerial quality is described as, “...the degree to which administrative systems such as procurement, human resources, and data management support the delivery of high-quality clinical care” (Moss and Garside 1995; Egger et al. 2005 as cited in
Bradley et al. (2010, 15). The authors note that administrative systems influence other organizational intermediate outcomes such as efficiency, access, and learning; consequently the authors assert that the contribution of a given managerial process to organizational performance must be assessed according to multiple intermediate outcome criteria.

Al-Qutob, Mawajdeh, Nawar, Saidi, and Raad (1998, 6) define management as, “...the set of all activities within the health care facilities through which the available human, physical, and financial resources are utilized efficiently to produce a given planned output.” Management refers to the means to achieve product or service quality. Rose (2005) outlines four components of quality that are worth noting: quality planning, quality control, quality assurance and quality improvement. The WHO Health System framework includes several components that are relevant for organizational performance such as service delivery, health workforce, information, and leadership/governance.

The conceptual and empirical literature notes that there are several legitimate and possible domains comprising management quality. For simplicity and comparability, management quality will be defined as a binary and qualitative category, involving presence or absence of a management mechanism that has been associated with high quality care. In outlining elements of quality management, Bradley et al. (2010) identify three important dimensions that aid in operationalizing this category: 1) quality assurance mechanisms; 2) information systems quality; and 3) governance quality – leadership.

**Additional Indicator Ranking Criteria**

In addition to the above considerations, the criteria used to select and rank good, better, best indicators for this category were: the level of accepted standardized processes engaged in by the program.

**Indicators for Measuring Management Quality**

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Indicator</th>
<th>Example</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>Existence of a monitoring and evaluation system</td>
<td>Marie Stops International: MSI Bolivia continuously monitors all organizational activities and conducts internal and external evaluations with an assessment tool that uses a standardized format for all of MSI’s regional programs; this creates transparency and accountability.</td>
<td>12.50%</td>
</tr>
<tr>
<td>Good</td>
<td>Existence of a performance review process</td>
<td>BlueStar Pilipinas: Franchisees fill out weekly performance reports, which include intrauterine device insertions and removals, number of injectables, number of condoms, number of pills (progestin-only pills and combined oral contraceptives), pap smears, new clients, and existing clients.</td>
<td>6.25%</td>
</tr>
<tr>
<td>Better</td>
<td>Existence of Audit</td>
<td>BlueStar Pilipinas: the Business Systems Audit is a scheduled visit and is conducted by a Field Team Member (FTM) two times per</td>
<td>1.25%</td>
</tr>
</tbody>
</table>

22 The denominator for this reporting frequency is 80 since we engaged in 80 full program extractions to collect data on this field.
<table>
<thead>
<tr>
<th>Ranking</th>
<th>Indicator</th>
<th>Example</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>year. This audit reviews performance records, stock keeping, payment history, and branding/promotion. For the performance records, the FTM reviews the service statistics that have been submitted via SMS to an automated computer program against the franchisees’ records. The FTM helps the franchisee correct any problems with entering the data;</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td>Existence of data management system</td>
<td><strong>Viva Sehat (formerly Razi Clinics):</strong> Razi has opted for iON (a branded on-demand cloud computing offering from Tata Consultancy Services) to digitize patient-care profiles that helps in easy access and more accurate treatment.</td>
<td>1.25%</td>
</tr>
<tr>
<td></td>
<td>Existence of standards and guidelines</td>
<td><strong>Hygeia Community Health Plan (HCHP):</strong> these are sets of standards, criteria or specifications required in the care of patients for particular medical conditions. They are quality management mechanisms designed to aid providers in making decisions about the most appropriate course of treatment for specific clinical cases. These guidelines help to standardize the care given to enrollees for specified conditions across board. Practice guidelines are developed from evidence based studies with the in-put of specialists and clinicians involved in patient care. These practice guidelines are subsequently reviewed with the providers before they are adopted and implemented. Practice guidelines have been developed for the more common conditions that afflict majority of the scheme members. Practice guidelines are presently available for malaria, hypertension and diabetes. Compliance with clinical practice guidelines is evaluated routinely during case management activities and case file audits.</td>
<td>1.25%</td>
</tr>
<tr>
<td>Ranking</td>
<td>Indicator</td>
<td>Example</td>
<td>Frequency</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>Best</td>
<td>International Organization for Standardization certification</td>
<td>Piramal E-swasthya: an ISO 9001 2008 certified company. Lifespring – as part of this focus on continual improvements, LifeSpring’s flagship hospital in Moula Ali was ISO 9001:2000 certified by TUV NORD in November 2007. LifeSpring’s hospitals are ISO-certified with strict clinical guidelines</td>
<td>2.50%</td>
</tr>
</tbody>
</table>

Total achieving: Good, better, or best 25%

Recommended Operational Definition
Management quality is a qualitative category, which makes it challenging to define precise metrics. Employing the four areas identified above (quality planning, control, assurance, and improvement), the following categories and indicators are recommended for the management quality field:

1. Quality assurance mechanisms:
   a. Audit
   b. Performance assessment
   c. Monitoring and evaluation system
   d. Monitoring system
   e. Independent evaluation
   f. Standards and guidelines
   g. Oversight bodies

2. Information systems quality
   a. Data management system
   b. Data collection and reporting tools

3. Governance quality
   a. Board of Governors
   b. Management committees

Management Quality: The procedures, systems, and processes the organization and/or program has implemented to strengthen quality in all aspects of operations and delivery.

Credibility, Feasibility and Comparability
The ranking of data in this category is based on quality standards in reporting. Data coded as good include simple references to management quality indicators with a brief description. Data coded as better includes mention of a key management quality indicator and should include a detailed description on how the management quality mechanism, procedure, or process operates. Data coded as best included reference to a standardized indicator such as International Organization for Standardization (ISO) certification with date information. International Standards ensure that products and services are safe, reliable and of good quality. Some of ISO’s most well-known standards are management system standards. They provide a model to follow when setting up and operating a
management system (International Organization for Standardization). For this category 18.75% of programs reported good data, 3.75% provided better reporting, and 2.50% best reporting.

Reporting in this category is highly feasible as collecting management quality data does not necessitate studies or evaluations but rather requires descriptions of the process, procedures, or mechanisms an organization has established. A program manager can easily collect this data. The data is difficult to compare across programs due its qualitative and non-standardized nature.

Credibility:  
Feasibility:  
Comparability: 

Measuring Human Resources Supply and Commitment of Staff

Literature Review

As stated by the WHO (2010, 24), “…the ability of a country to meet its health goals depends largely on the knowledge, skills, motivation and deployment of the people responsible for organizing and delivering health services.” To strengthen the performance of health systems, actions must be taken for assessing and strengthening the recruitment, retention, and productivity of health workers (WHO 2010). Qualified and motivated human resources are integral for effective health services provision (WHO 2006). Human resources in the health field include clinical staff, such as nurses, physicians, dentists, and pharmacists, as well as management and support staff that may not directly deliver services but are vital to the performance of health systems (WHO 2010).

In the health systems literature, human resources are described as a core component or building block of health systems (WHO 2010). De Savigny et al. (2009) note that human resources include integrated training and incentives for retention or remote area deployment, in addition to quality improvement and performance management23. Within the CHMI Plus framework, a distinction has been made between human resources supply and human resources management, the latter of which is being captured under management quality. Bradley et al. (2010) describes access to a supply of trained human resources as a component of sustainability, which includes a, “…robust connection with health workforce educational pipeline.” This approach includes two components: 1) access to a trained workforce, and 2) efforts to ensure the workforce is trained. Bradley et al.’s (2010) framework suggests that both recruitment and training are integral aspects of the human resource supply.

In Bradley et al.’s (2010) literature review of health care organizational performance, the authors consider commitment of staff as an aspect of sustainability. Commitment of staff includes several components such as commitment to the organization, provider satisfaction with work, provider’s perception of safety, staff satisfaction with work, staff support and motivation, work climate, and work-related stress (Bradley et al. 2010). There are significant costs associated with staff members leaving an organization, which includes long-term costs for hiring and training new staff (West and Dawson 2011). Indeed, staff training and turnover carry significant costs, both financial and non-financial, including loss of intellectual property (Karamchandani 2009). To reduce the costs of avoidable staff turnover, it is helpful to understand factors influencing aspects of human resources. Staff members working in

---

23 Quality improvement and performance management are captured under management quality in the CHMI Plus framework.
supportive and effective teams are less likely to leave (West and Dawson 2011), and financial incentives may be used to promote staff retention among top performers (Karamchandani 2009). Shen, Cox, and McBride (2004) identify the following factors that affect retention: a fair salary (compared to that of peers), benefit packages composed of financial and non-financial incentives, flexibility in working time, career development, and working environment, including teamwork, and relationships with managers.

The ability to recruit, train, and retain trained staff is an important aspect of health services delivery as poor performance of health providers can lead to inaccessible and ineffective health services provision (WHO 2006). While Bradley et al. (2010) have disaggregated components of human resources under different domains and dimensions, given the similarities and interconnections between these domains and dimensions, we have combined the various components into a category called human resource supply and commitment of staff.

Additional Indicator Ranking Criteria
In addition to the above considerations, the criteria used to select and rank good, better, best indicators for this category were: the level of detail and description, and the inclusion of results from a process.

Indicators for Measuring Human Resources Supply and Commitment of Staff

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Indicator</th>
<th>Example</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>Provides training, recruitment, and/or retention programs</td>
<td>Community Health Africa Trust: Train family planning community based distributors</td>
<td>7.14%</td>
</tr>
<tr>
<td>Better</td>
<td>Describes efforts to promote recruitment and/or retention; description of training programs</td>
<td>CARE Hospitals: In order to have a regular flow of staff, CARE offers various education programs. The programs consist of Post Graduate Diploma Programme in Community Cardiology (PGDPCC), Physician Assistant, MSc in Hospital Administration for Nursing Graduates, Post-Graduate Diploma Course in Cardiovascular Nursing and Clinical Research Training. Among the total workforce, 50 percent are nursing staff and CARE has tied up with many other colleges to ensure a sufficient supply of staff. CARE has defined the career path for all cadres.</td>
<td>21.43%</td>
</tr>
</tbody>
</table>

The denominator for this reporting frequency is 84 since we engaged in 80 full program extractions to collect data on this field, in addition to four targeted extractions to collect more information on this and other low-reporting categories.
| Turnover or retention rate per reporting period | **Aravind Eye Care System:**
Per 2007: the median turnover rate was 7.2%. | 4.76% |
| --- | --- | --- |
| Describes staff satisfaction and/or factors detracting from retention and/or recruitment | **Sun Quality Health Network, Myanmar:**
“Personal reasons” is the most common reason cited for leaving the network. Others have disassociated because they have died, migrated, suffered ill health, or have been dismissed. | 4.76% |
| Qualitative and quantitative description of provided recruitment, training, and/or retention programs and impact of programs | **Hygeia Community Health Plan (HCHP):**
The workshops provided by HCHP typically involve all categories of healthcare workers in the hospitals i.e. Doctors, Nurses, Administrators, Laboratory and Pharmacy personnel. The choice of training topics is guided by the utilization statistics. The workshops focus on the most prevalent conditions to ensure that evidence-based care and quality services are dispensed by the healthcare providers. This approach stimulates an integrated or systems view of the medical condition, in which all weak links in the service chain are addressed and strengthened. | 5.95% |

**Total achieving: Good, better, or best** 44%

**Recommendations/Operational Definition**

**Human Resource Supply and Commitment of Staff:** Description of provided programs that seek to promote recruitment or retention; description of training programs; turnover or retention rate per reporting period; and/or description of staff satisfaction and/or factors detracting from retention and recruitment.

**Credibility, Feasibility, and Comparability**
In the human resource supply and commitment of staff category, we have grouped together recruitment, training, and retention due to their overlap and similarity. While the literature on human resources describes the importance of strengthening recruitment, training, and retention of staff in order to ensure strong health performance, these aspects have been placed under different assessment fields. Although our approach differs from the literature, components of human resource supply and commitment of staff are well described in the literature, and retention in particular is a very credible measure. This definition is considered to be highly feasible given that a general description of provided recruitment, training, and/or retention programs should not require undo effort and additional study measures. Retention or turnover rates are also considered feasible given that information on the number of employees should be well documented. The comparability of this field is low due to the descriptive nature of the measure and the wide variety of types of health programs and organizations.
and the employees they are aiming to recruit, train, and retain. For this category 7.14% of programs reported good data, 30.95 % provided better reporting, and 5.95% best reporting.

Credibility:  
Feasibility:  
Comparability:  

Measuring Political Support

Literature Review

The political support category stems from Bradley et al.’s (2006) framework where it is placed under the field of sustainability. Here, sustainability is described as the program’s “...ability to continue delivering needed and valued services” (Bradley et al. 2006, 5). This perspective includes the notion from open system theories that organizations and programs are both shaped by their environment, and as actors themselves, shape their context (Scott 2004). In this view, organizations are embedded in an environment and part of system that includes political, social, economic and cultural institutions (Scott 2004). This approach considers the impact of the given context or environment on the continued operations of the program, and acknowledges that the programs’ interaction with broader political systems can strongly influence its current and future performance (Bradley et al. 2006).

Political support is a significant component of sustainability. Olsen (1995) describes sustainability as involving several factors, including sustained political support from government officials. As well, in describing the sustainable development of primary health care services, the Canadian Public Health Association (1990, 4) notes that political sustainability involves, “...development and maintenance of the political will necessary to sustain a major policy direction.” Thus, this category considers the relationships fostered between the program and its political environment, including forms of government at national, regional, and local levels. This is a difficult category to develop metrics for given the widely varying political contexts and governmental structures that exist not only in different countries but also in different regions and communities.

Ideally, one would have a strong understanding of the political climate and context when analyzing such metrics. We have focused this category on the relationship between programs and government entities from the local to national levels in order to keep the reporting for this complex field both focused and feasible. The ability to succinctly yet sufficiently describe the program’s engagement with national, regional, and local government entities can provide a useful window into the ability of the program to gain necessary political support and legitimacy to continue carrying out its activities.

Additional Indicator Ranking Criteria

In addition to the above considerations, the criteria used to select and rank good, better, best indicators for this category were: the level of detail and description, and the inclusion of a reporting period.
<table>
<thead>
<tr>
<th>Ranking</th>
<th>Indicator</th>
<th>Example</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>Listing of a partnership with a local, regional, or national government entity</td>
<td><strong>Learning about Living</strong>: Operational partners include the Nigerian Federal Ministry of Education (HIV/AIDS Unit), and the Nigerian Federal Ministry of Health.</td>
<td>28.57%</td>
</tr>
<tr>
<td>Better</td>
<td>Listing of a partnership with a local, regional, or national government entity per reporting period</td>
<td>N/A</td>
<td>0%</td>
</tr>
<tr>
<td>Best</td>
<td>Qualitative evidence of a relationship or partnership with a local, regional, or national government entity (including financial or technical support, authorization or approval of activities, successful advocacy, and/or training of government officials)</td>
<td><strong>Marie Stopes Bolivia</strong>: This program has negotiated agreements with the national Ministry of Health and with local departments and municipalities that enable them to use mobile units to provide services at municipal health posts or health centers. The municipality schedules a visit from MSI Bolivia, organizes potential clients and provides a place for the services. <strong>Population and Community Development Association (PDA)</strong>: In cooperation with the Ministry of Interior, PDA trained government officers, district officers, governors, and community leaders, such as monks, to inform the wider community about AIDS.</td>
<td>8.33%</td>
</tr>
</tbody>
</table>

**Total achieving: Good, better, or best** 39.28%

**Recommended Operational Definition**

**Political Support**: Qualitative evidence of a relationship or partnership with a local, regional, or national government entity per reporting period (including financial or technical support, authorization or approval of activities, successful advocacy, and/or training of government officials).

---

25 The denominator for this reporting frequency is 84 since we engaged in 80 full program extractions to collect data on this field, in addition to four targeted extractions to collect more information on this and other low-reporting categories.
Credibility, Feasibility, and Comparability

The credibility of this recommended definition has been assessed above in the table displaying indicator rankings. While the literature agrees that political support is an important aspect in determining sustainability, it is not often well defined, suggesting that this definition may not have strong credibility. However, it would be very feasible for programs to briefly describe their relationship with local, regional, and/or national government entities as well as providing a time period for that relationship. This approach offers a richer understanding of the political support for the program and allows for greater credibility of this measure beyond simply listing a government entity as a partner. In terms of comparability, it is very difficult to compare the political support of programs operating in different contexts with different histories and political structures. Still, developing an understanding of the relationship dynamics between programs and government entities can provide insight into the ability of these programs to generate necessary political support for their current and future activities. For this category 28.57% of programs reported good data, 8.33% provided better reporting, and 2.38% best reporting.

<table>
<thead>
<tr>
<th>Credibility:</th>
<th>Feasibility:</th>
<th>Comparability:</th>
</tr>
</thead>
</table>


Literature Review
The CHMI Plus framework adopts the IRIS framework indicators for financial performance. IRIS developed these indicators based upon both the Generally Accepted Accounting Principles (GAAP) and the International Financial Reporting Standards (IFRS). The CHMI Plus framework groups these indicators into the following fields: income statement, balance sheet, cash flows, and ratios, concepts, and calculations. Many programs collect financial data without making it publicly accessible. Data may be restricted to an internal or limited audience (e.g. potential investors). Our extractions and recommendations are based upon publicly available data.

Extraction Review
From our extraction of publicly available information on 84 programs, we have been able to collect data on the fields below:

- Income Statement: 51.19%
- Balance Sheet: 69.05%
- Cash Flow: 9.52%
- Ratios, Concepts, and Calculations: 10.71%

Much of the data collected in these fields are qualitative in nature, referencing primary/additional sources of funding or listing funders. The percentage of programs reporting revenues, budgets, or donor funding amounts is 33.33%.

---

26 Extractions focusing on gathering data for all categories occurred for a subset of 80 programs from CHMI’s database. In an attempt to gather more data for under reporting categories, extractions focusing only on certain categories, including income statement, balance sheet, cash flow, ratios, concepts and calculations, took place for an additional four programs.
Good practices
We recommend the following as an acceptable minimum level of data related to financial performance be provided. Based on the financial reporting literature and best practice, the following five categories have been identified as the minimum acceptable level: 1) revenues/budget/donor funding, 2) country of registration and legal status, 3) currency, 4) tax status, and 5) number of employees.

Best reporting in Extraction: Operation ASHA
Out of the 84 programs from the sample extractions, Operation ASHA provides best reported publicly available data. The format, quantity, and content of the extracted indicators are useful for analysis and comparability. Operation ASHA provides program-specific data, allowing viewers to examine the program’s financials. Each extracted indicator provides three years of annualized data, allowing viewers to examine possible trends and compare performance across programs. While the data made publicly available from Operation ASHA fails to include several financial measures of performance, it provides sufficient information for computing these indicators.

Credibility, Feasibility, and Comparability
Most programs did not provide significant information on financial performance. Common problems include: no data or poorly reported data, data aggregated at an organizational level, and financials audited at an organizational level. Financial data was unavailable, at least publicly, in the sample extractions. Where present, the financial indicators were often presented in a format that limited their comparability. Many programs aggregate financial data at the organizational level, making it difficult to discern the program’s financial performance within the organization.

Another significant challenge in reporting financial performance data is it requires a requisite level of technical knowledge regarding accounting and finance. While the standards developed by IRIS are extremely robust and credible, the ability to report on these measures mandates that program managers possess the ability to understand and collect such data. Our sample extractions revealed that financial performance data was poorly reported; however, some programs may still be collecting this data without making it publicly available.

<table>
<thead>
<tr>
<th>Program</th>
<th>2508 - Operation ASHA</th>
</tr>
</thead>
</table>
| Income Sheet | Amounts in Rupees (INR):  
Earned Revenue 2010-11: Portion of 20335786.88 (Membership, grant and donation)  
Earned Revenue 2009-10: Portion of 8001608.79 (Membership, grant and donation)  
Earned Revenue 2007-08: Portion of 2927808 (Membership, grant and donation)  
COGS 2010-11: 19652652.61  
COGS 2009-10: 7823217.31  
COGS 2007-08: 2534223.82  
Depreciation 2010-11: 374287  
Net Income 2010-11: 401251.2  
Net Income 2009-10: 57522.91  
Net income 2007-08: -309590.53 |
Program 2508 - Operation ASHA

<table>
<thead>
<tr>
<th>Balance Sheet</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary funding source: Government</td>
<td></td>
</tr>
<tr>
<td>Additional source of funding: Donor</td>
<td></td>
</tr>
<tr>
<td>Amounts in Rupees (INR).</td>
<td></td>
</tr>
<tr>
<td>Current assets 2010-11: 12784003.35</td>
<td></td>
</tr>
<tr>
<td>Current assets 2009-10: 2821290.38</td>
<td></td>
</tr>
<tr>
<td>Current assets 2007-08: 35385.84</td>
<td></td>
</tr>
<tr>
<td>Fixed assets 2010-11: 2998624.5</td>
<td></td>
</tr>
<tr>
<td>Fixed assets 2009-10: 1174391.5</td>
<td></td>
</tr>
<tr>
<td>Fixed assets 2007-08: 290404</td>
<td></td>
</tr>
<tr>
<td>Total assets 2010-11: 15782627.85</td>
<td></td>
</tr>
<tr>
<td>Total assets 2009-10: 4073240</td>
<td></td>
</tr>
<tr>
<td>Total assets 2007-08: 325790.84</td>
<td></td>
</tr>
<tr>
<td>Accounts payable 2010-11 (&quot;expenses payable&quot;): 871934.77</td>
<td></td>
</tr>
<tr>
<td>Accounts payable 2009-10 (&quot;expenses payable&quot;): 622240</td>
<td></td>
</tr>
<tr>
<td>Accounts payable 2007-08 (&quot;expenses payable&quot;): 12500</td>
<td></td>
</tr>
<tr>
<td>Total liabilities 2010-11: 15782627.85</td>
<td></td>
</tr>
<tr>
<td>Total liabilities 2009-10: 4073240</td>
<td></td>
</tr>
<tr>
<td>Total liabilities 2007-08: 325790.84</td>
<td></td>
</tr>
</tbody>
</table>

Ratios, Concepts, and Calculations

Under the Operation ASHA’s model (where the treatment per patient-approximately $30-equals the total sum granted by the government), every treatment center is financially self-sustainable after two years.

Discussion

This report developed a performance measurement framework that is based on available information on innovative health programs in low and middle-income countries and the literature on organizational performance. We collected detailed information on 80 programs from publicly available sources to supplement the Center for Health Market Innovations (CHMI) dataset and used this to define best practice in performance reporting. The process involved developing clear definitions for key indicators of performance, providing examples, describing the frequency of use, and assessing their credibility, feasibility, and comparability. The extracted programs encompass a broad number of health areas, such as HIV/AIDS, cancer, malaria, maternal health, and heart disease. Also, many of the programs employ a variety of approaches, such as microinsurance, mHealth, mobile clinics, and franchising. Building from the literature, we have designed a framework with performance metrics that is relevant to a diverse set of programs and able to effectively capture their results. Developing aggregate indicators that cut across thematic areas is challenging, and this work will require feedback from third-party evaluators, program managers, and investors to assess its usability and ensure its relevance.

Empirically, in developing and assessing the CHMI Plus framework, we have focused on gathering data on 80 programs in the CHMI database. We have supplemented information in the database by collecting publicly available data from journal articles, reports, and reputable organization and news websites. In doing so, we have developed a user guide to assist third party data collectors and program managers in gathering data for the CHMI database. We have found the publicly available data to be rich and
informative, providing not only useful metrics, but also descriptive information that lends valuable context to the health service offerings.²⁷

The following table summarizes the frequency of reporting “good-better-best” indicators in the sample, as well as our assessment of the indicators’ credibility, feasibility, and comparability (indicators are rated from 1 to 3, with 3 indicating a high score).

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Achieving &quot;good-better-best&quot;</th>
<th>Credibility</th>
<th>Feasibility</th>
<th>Comparability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population coverage</td>
<td>11%</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Output</td>
<td>61%</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Health outcome</td>
<td>46%</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Affordability</td>
<td>37%</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Physical access</td>
<td>20%</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Pro-poor targeting</td>
<td>16%</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Clinical quality</td>
<td>19%</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>User satisfaction</td>
<td>28%</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Management quality</td>
<td>25%</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Human resources</td>
<td>44%</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Political support</td>
<td>39%</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Financial data</td>
<td>10% to 70%</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Credibility, Feasibility, and Comparability

In addition to describing best practice in performance measurement, we have assessed the credibility, feasibility, and comparability of each dimension of performance. While credibility, feasibility, and comparability of fields are relevant for investors, program managers, and researchers, they may be of greater interest to certain stakeholders as compared to others. Our aim was to develop category definitions that reflected an optimal balance of these three criteria.

Credibility was assessed based on the health services, program evaluation, and management literature. If there was a robust conceptual and empirical literature informing the category, and a high level of agreement on how a category should be measured, credibility received a high rating. The credibility ranking indicates the degree to which indicators reported reflect what is recommended in the literature; it also provides an assessment of our operational definition. The credibility rating is most relevant for investors and researchers.

- Based on these assessment criteria the following fields received the highest credibility rating: financial management, health output, and population coverage.
- The following field received the lowest credibility rating: political support.

Feasibility was assessed based on how practical it would be for program managers to collect data on a particular dimension of performance. Certain categories demand a significant amount of resources for

²⁷ Two program profiles produced from data gathered using the CHMI Plus framework can be found in Appendix C.
data collection, while others are readily available or routinely collected by most organizations. Feasibility is especially relevant for program managers, who are responsible for data collection.

- The following fields received the highest feasibility rating: health output, human resources supply/commitment of staff, and management quality.
- The following received the lowest feasibility rating: affordability, clinical quality, financial management, health outcome and impact, physical access/service availability, political support, and population coverage.

**Comparability** was rated based on the ability to compare indicators in a category across programs. Comparability is important for researchers, and investors that are interested in benchmarking.

- The following field received the highest comparability rating: financial management.
- The following fields received the lowest comparability rating: affordability, human resource supply/commitment of staff, management quality, political support, and pro-poor targeting.

**Strengths**

This report combines an extensive review of the health services literature on performance measurement with the current reporting practices of a large sample of programs. Previous studies, like the systematic review by Bradley et al. (2010), focused on the peer reviewed literature, which includes many third party and academic evaluations that are not feasible for most organizations in low-resource settings. This study could compare up to 80 different examples of how a given dimension of performance is reported. Many proposed frameworks do not consider how to compare across a large number of different disease areas and strategies. When recommending categories and indicators, this report systematically considers dimensions of credibility, feasibility, and comparability, which is often a theoretical concern, but precedence is generally given to credibility, which may limit the applicability of existing frameworks. Since the CHMI database includes a wide range of programs and strategies, and includes emerging areas like mHealth, the proposed framework identifies principles and performance metrics that are widely applicable. We feel this is a strong first step towards building more standard approaches to reporting that can build confidence for investors, increase the learning for program managers, and allow researchers to highlight emerging practices that show promise.

**Limitations**

We note several limitations to our approach. First, in the extraction process, we were limited to accessing publicly available data from websites, reports, and articles. We often did not have access to internal program performance reports that would likely provide greater information and richer detail for our dataset. Thus, our extractions reflect the data that is available to third-party data collectors rather than the data available to program managers with access to internal reports.

Second, we did not assess the credibility of the sources used to gather data. We focused on using the program’s websites, as well as the reputable websites of news organizations and funding agencies, journal articles, and organizational reports. The credibility of sources is an important aspect in determining the quality of the reported data, and in our following discussion about evaluating indicators, we note that the quality of the data source should be included when assessing program data, as recommended by CHMI.

Third, our focus in producing a hierarchy of good, better, best indicators and a recommended definition is to provide a simple and feasible indicator that is also credible and comparable. In this process, we have found that some data apply to several categories, and we have tried to place this information in a single field in order to avoid duplication and overlap in collected data. In creating these boundaries,
however, we acknowledge that some overlap may still occur; these categories cannot cover all possible areas of reported data. They do, however, aim to cover the most pertinent reported fields based on the health services literature and our review of the publicly reported data on these programs.

Fourth, there continue to be challenges in distinguishing between aggregate organizations and individual programs, such that some data is reported at the organizational level rather than at the program level. In determining whether to include data on related activities offered by the program’s organization, extractors use a value chain approach. This involved including activities that were integral in contributing to the operations of the main innovative program described in the CHMI database.

**Conclusion and Next Steps**

In this report, we have reviewed best practices in performance reporting and developed a hierarchy of validity of performance measures. This process involved supplementing data in the CHMI database for 80 programs by extracting data from publicly available sources, followed by a literature review focused on refining framework categories. This approach resulted in a description of good, better, best indicators for each framework category, as well as operational definitions based on an assessment of the optimal balance of credibility, feasibility, and comparability. We have developed a user guide incorporating these definitions; this guide is intended to assist program managers and third-party data collectors in gathering data on CHMI programs for the CHMI database. Our next report will develop composite measures linking inputs and outcomes based on existing data and we will also begin to consider approaches to the analysis of strategies among the best-documented programs. These are key steps needed to develop a comprehensive framework for performance measurement that will be acceptable to key stakeholders and promote effective innovations to address the health needs of the poor.
References


British Medical Bulletin 1-17.


APPENDIX A
CENTER FOR HEALTH MARKET INNOVATIONS

CHMI Plus Framework

User Guide
User Guide for CHMI Extractions

Introduction
The purpose of this user guide is to provide guidance on how you as a program manager can input relevant data from your program’s performance into the CHMI Plus framework. The guide offers guidance on the protocol that should be used to input the data into the various fields, definitions of the indicators, and concrete examples of how data and results can be reported for the different categories and indicators.

This user guide is a living document. It will be updated periodically as CHMI and health services delivery organizations continue to report on a body of specific knowledge and experience in the health services field.

Guidance for entering data into CHMI database

References
List all references used in the extraction except information directly sourced from the CHMI database and/or website. Citations should be listed in the "References" column.

Data source
Data should be derived from sources that have documented the program’s performance over time. Data sources for indicators can be administrative, primary, or secondary (Treasury Board of Canada Secretariat):

- **administrative data** - information that is already being collected in program or initiative files or databases, or could be collected from existing program sources;
- **primary data collection** - information that needs to be collected through specialized data collection exercises such as focus groups, expert panels or surveys; and
- **secondary data** - data that have been collected for other purposes, but which could also be used in this context, such as national statistics on health or economic status, for example.

Secondary data must be used with caution as such data will have been gathered with other organization goals or agendas in mind. Secondary data means using someone else’s data to report progress and success in moving toward your own desired outcomes. As a program manager, it is important to consider whether such secondary data is valid and reliable. Anecdotal information that contains opinions of individuals on the program’s performance should not be included in the database (Kusek and Rist 2004).

Indicators
The CHMI Plus framework in the tables below includes a list of indicators that are meant to guide program managers on providing data on program results. The indicators included below are not exhaustive and include only sample of relevant indicators. Program managers are advised to report results based on their relevant performance indicators.

---

28 Program managers are advised to report results based on their relevant performance indicators.

29 This condensed version of the user guide does not contain the sample list of indicators.
An indicator is a quantitative or qualitative factor or variable that provides a simple and reliable basis for assessing achievement, change or performance in program results. Indicators are the quantitative or qualitative variables that provide a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of an organization against the stated outcome.

Indicators help achieve the following (Bastia 2000):

- **Measurement of change**: Indicators are designed to measure changes. A number that conveys information but does not give information in relation to changes is a statistic.

- **Direction**: Indicators are useful for pointing the direction of change, whether this is positive or negative, whether the situation is improving or worsening.

- **Measurement of change over time**: Indicators are designed to measure change over time.

Quantitative indicators are often expressed as the number or percent (proportion or rate) of something. Program managers should include both forms. The number of successes (or failures) in itself does not indicate the rate of success (or failure) - what was not achieved. The percent by itself does not indicate the size of the success. Population coverage, for example, is only useful if a numerator and denominator are provided:

> The mountainous rainforest in which the ABC hospital is located is home to target population of roughly 60,000 people. When XYZ program was first introduced in 2010, approximately 27.6% (16,560 clients) of this population made use of ABC's medical services. After the introduction of XYZ, by 2011, the medical care utilization had doubled to approximately 55.2% (33,120 clients) of the population.

As best practice, data that includes rates or proportions should provide information on the numerator and denominator.

Qualitative indicators provide insights into people’s judgments, opinions, perceptions and attitudes towards a given situation or subject. They can include changes in sensitivity, satisfaction, influence, awareness, understanding, attitudes, quality, perception, dialogue or sense of well-being (UNDP 2009).

The following is an example of a qualitative indicator:

> Existence of mechanisms, such as surveys, for obtaining timely client input on the existence of appropriate, timely and effective access to health services.

**Timely data and reporting period**

The CHMI framework requests data per reporting period. The reporting period refers to “the time from the reported Report Start Date to the Report End date” (GIIN 2012). The data reported annually by calendar year will be classified by the year for which they were reported. If annual data is not available, program managers should provide data for the months available.

As best practice, programs should report data on an annual basis. Data can date back to five years for inclusion in the framework. Data older than five years should not be included in reporting results. Data points that are less than or more than one year should be annualized prior to being entered into the database.
**Annualizing data**
The data reported annually by calendar year will be classified by the year for which they are reported. To annualize is to take an item measured for a period other than one year and compute what it would be for a single year. Although it is possible to annualize numbers for periods longer than one year, usually the items to annualize are for periods of 12 months or less. For example, if program managers annualize a production level of 30,000 units per quarter, it becomes 120,000 units a year. Or if a program annualizes a salary of $3,000 a month, it becomes $36,000 a year. The effort to annualize data has drawbacks: it does not take into account seasonal variations or any volatility in a number during the year. Note that to annualize a figure is not to compound it, e.g., interest on an investment posted each month would not be assumed to be reinvested. Also, to annualize an interest rate on a loan is not the same as computing its annual percentage rate, which is often based on many factors (Investor Glossary).

The data reported annually by calendar year will be classified by the year for which they were reported. The data reported by fiscal year will be classified by calendar years, with data from fiscal years ending on or before June 30 placed in the previous calendar year, and those ending after June 30 applied to the following calendar year. For example, data reported for a fiscal year ending on May 31, 2010 would be classified as data from 2009, while data from a fiscal year ending on August 31, 2010 would be classified as data from 2010.

For data that are reported for shorter time periods, the following process should be used to annualize these data:

- For “balance” or “instant” metrics (e.g. total assets or number of employees) any reported number will be applied to the date it is reported.
- For “duration” or “period” metrics (e.g. earned revenue or payments to smallholder farmers) the data should be annualized by adding periods together to create a calendar year:
- If twelve months of consecutive data are available, but do not align with a calendar year, then the same principle that governs the classification of fiscal years should be applied (periods ending on or before June 30 were classified in the previous calendar year, and those ending after June 30 were classified in the next calendar year) (IRIS).

In cases when a year’s worth of data is not available, consistent performance should be projected to complete a year of data. For example, an organization that reported $3 million in revenue for 9 months would have been projected to achieve $4 million in revenue for the year.

**Baselines**
As best practice, program managers should provide information on their baselines. The performance baselines establish a starting point from which to later monitor and evaluate results. Baseline data is critical for evaluating program performance. They are useful in providing information that may be used later to provide a comparison for assessing program’s outcomes or impacts (World Bank 2010).

If baseline information is not available, program managers are encouraged to provide data that demonstrates how program performance has changed. For example, the program’s performance for the current year can be compared to the year before to assess whether performance according to different indicators has improved or worsened.

**Targets/Benchmarks**
Targets can be selected by examining baseline indicator levels and desired levels of improvement. Targets are the quantifiable levels of the indicators that a country, society, or organization wants to
achieve by a given time (Kusek and Rist 2004). For example, one target could be “80% vaccination coverage for measles in target population.”

Benchmarks are standards against which a result is measured. Benchmarking is a systematic process for identifying and implementing best practices (Business Performance Improvement Resource). While setting targets and benchmarks is considered best practice, much of the reporting in the CHMI database did not include targets and/or benchmarks.

Targets can be developed for many of the CHMI categories including outcomes, outputs, population coverage, affordability, etc. While no separate field has been created to capture targets/benchmarks in the framework, program managers are encouraged, where possible, to provide information on their targets/benchmarks.

### Definitions

**Innovations**

We adopted Weberg et al.’s definition of innovation in health care, which emphasizes the impact of the innovation on the market or population: "Innovation is something new or perceived new by the population experiencing the innovation, that has the potential to drive change and redefine healthcare's economic and/or social potential” (Bhattacharyya, Khor, McGahan, Dunne, Daar, and Singer 2010, 2). The "newness" in an innovation can be achieved by "recombining old ideas in a new way, creating a new process or product, using a process from another industry in one that has not used that process, or reordering an organization in a new and different way" (Bhattacharyya et al. 2010, 2). Business models consist of four components: i) a product or service; ii) managers that bring together a set of resources required to deliver the product or service; iii) processes where employees and resources work together to repeatedly generate the product or service; and iv) a profit formula to ensure that the costs of the resources and processes are covered (Hwang and Christensen 2008).

The box below provides the definitions that CHMI uses to define different types of innovations. Program Managers should assess whether and how their programs have achieved different types of innovations according to the five categories below.

### CHMI Innovation Types

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Organizing Delivery</strong></td>
<td>Programs that reduce fragmentation and informality of health care delivery</td>
</tr>
<tr>
<td><strong>Financing Care</strong></td>
<td>Programs that mobilize funds and give purchasing power to the poor</td>
</tr>
<tr>
<td><strong>Regulating Performance</strong></td>
<td>Programs that set standards, enforce and promote quality care</td>
</tr>
<tr>
<td><strong>Changing Behaviors</strong></td>
<td>Programs that educate and train consumers and providers to seek and deliver better care</td>
</tr>
<tr>
<td><strong>Enhancing Process</strong></td>
<td>Programs that apply new technologies and operational processes to improve quality, access or cost</td>
</tr>
</tbody>
</table>
**CHMI Plus Framework**

The tables below describe the different categories that comprise the CHMI Plus framework, their definitions and where relevant some indicators for the different fields. It is important to note that the list of indicators below is not exhaustive. Program managers should report on results using the relevant indicators that pertain to their programs.

**Table 1: CHMI Plus Framework (Descriptive Fields)**

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Program Profile</strong>:</td>
<td>This field should contain a description of the following aspects of the program:</td>
</tr>
<tr>
<td>1. Summary</td>
<td>150 word write-up on the program.</td>
</tr>
<tr>
<td>2. Implementing organization</td>
<td></td>
</tr>
<tr>
<td>3. Health focus</td>
<td></td>
</tr>
<tr>
<td>4. Types of products/services offered</td>
<td></td>
</tr>
<tr>
<td>5. Country registered and legal status</td>
<td></td>
</tr>
<tr>
<td>6. Country of operation</td>
<td></td>
</tr>
<tr>
<td>7. Stage</td>
<td></td>
</tr>
<tr>
<td>8. Year launched</td>
<td></td>
</tr>
<tr>
<td>9. Number of facilities</td>
<td></td>
</tr>
<tr>
<td>10. Number of employees</td>
<td></td>
</tr>
<tr>
<td>11. Target population</td>
<td></td>
</tr>
<tr>
<td><strong>Problem</strong>:</td>
<td>This field describes in 100 characters or less the problem that the program is trying to address. This can include the rationale and/or justification for the program.</td>
</tr>
<tr>
<td><strong>Goal</strong>:</td>
<td>This field answers <em>WHAT</em> the program is trying to achieve. What is the aim or intention of the program?</td>
</tr>
<tr>
<td><strong>Process</strong>:</td>
<td>This field provides information on the <em>HOW</em>. How does the program achieve its goal(s)? This field should outline the processes and steps that are used to deliver the program’s product and/or services and the relationships between them.</td>
</tr>
<tr>
<td><strong>Challenges</strong>:</td>
<td>This field should describe the obstacles your program faces in delivering its product or service.</td>
</tr>
<tr>
<td><strong>Opportunities</strong>:</td>
<td>This field should describe any opportunities your program has discovered and plans to leverage</td>
</tr>
<tr>
<td><strong>Strategic Planning</strong>:</td>
<td>Describe how your program sets its plans for identifying and achieving future goals including scaling-up or plans for growth. This section should also include plans for engaging in activities to obtain resources and assigning responsibilities to attain these goals. This section should also provide information on the future plans of the program.</td>
</tr>
<tr>
<td><strong>Strategies</strong>:</td>
<td>This section should describe innovative strategies being used by the program</td>
</tr>
<tr>
<td><strong>Table 2: CHMI Plus Framework (Indicator Fields)</strong></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Population Coverage</strong>: Volume of clients served as a percentage of a defined target population per reporting period.</td>
<td></td>
</tr>
<tr>
<td><strong>Health Output</strong>: Quantitative evidence about the number of health services/products provided and/or clients served / trained per reporting period.</td>
<td></td>
</tr>
<tr>
<td><strong>Health Outcome and Impact</strong>: Quantitative evidence of intermediate or long-term health outcome evidenced by changes in learning, actions, or health status of clients/target population per reporting period.</td>
<td></td>
</tr>
<tr>
<td><strong>Affordability</strong>: Quantitative evidence about the price of services and/or product in comparison to the average cost of accessing similar services in the local context or as a proportion of income.</td>
<td></td>
</tr>
<tr>
<td><strong>Physical Access and Service Availability</strong>: Quantitative evidence about the ability of clients/patients to access health services at the needed place and time per reporting period. Measures include: number of facilities, providers or hospital beds per segment of population or target population, distance to facility, staff absenteeism, medical supply stock-outs, and hours of operation.</td>
<td></td>
</tr>
<tr>
<td><strong>Pro-Poor Targeting</strong>: The extent to which the program’s clients are disadvantaged in terms of socio-economic status and the criteria used to determine and target the poor. This includes reference to whether the targeting is being done in a predominately poor area or is targeting the most disadvantaged group within a population.</td>
<td></td>
</tr>
<tr>
<td><strong>Clinical Quality</strong>: Quantitative evidence of providing safe, effective, evidence-based care, compared to other providers of similar services, or demonstrates change over time.</td>
<td></td>
</tr>
<tr>
<td><strong>User Satisfaction</strong>: Quantitative or qualitative evidence that is collected using a systematic methodology and reflects the client’s perception on the quality of services provided.</td>
<td></td>
</tr>
<tr>
<td><strong>Management Quality</strong>: The procedures, systems, and processes the organization and/or program has implemented to strengthen quality in all aspects of operations and delivery.</td>
<td></td>
</tr>
<tr>
<td><strong>Cost-to-Service Ratios</strong>: The cost of providing a service and/or product per client.</td>
<td></td>
</tr>
<tr>
<td><strong>Staff-to-Service Ratios</strong>: The number of staff or devices allocated per client and/or procedure.</td>
<td></td>
</tr>
<tr>
<td><strong>Patient or Procedure Volume per Time Period</strong>: The patient volume per time period or the procedure volume per time period.</td>
<td></td>
</tr>
<tr>
<td><strong>Commitment of Staff / Human Resources Supply</strong>: Description of provided programs that seek to promote recruitment or retention; description of training programs; turnover or retention rate per reporting period; and/or description of staff satisfaction and/or factors detracting from retention and recruitment.</td>
<td></td>
</tr>
<tr>
<td><strong>Operational and Technical Partners</strong>: List of partnerships with different organizations that provide technical and operational support to the program. Technical partners typically refer to partners that provide guidance and advice on both a conceptual level and in implementing the project. Operational partners are typically those that play a role in implementing the program. Technological partners can be either or both of these categories.</td>
<td></td>
</tr>
<tr>
<td><strong>Political Support</strong>: Qualitative evidence of a relationship or partnership with a local, regional, or national government entity per reporting period (including financial or technical support, authorization or approval of activities, successful advocacy, and/or training of government officials).</td>
<td></td>
</tr>
<tr>
<td>Financial Statement: Financial performance is assessed by giving a summary of how the business incurs its revenues and expenses through both operating and non-operating activities. It also shows the net profit or loss incurred over a specific accounting period, typically over a fiscal quarter or year (Investopedia n.d.)</td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Balance Sheet:</strong> A financial statement that summarizes a company's assets, liabilities and shareholders' equity at a specific point in time. These three balance sheet segments give investors an idea as to what the company owns and owes, as well as the amount invested by the shareholders (Investopedia n.d.).</td>
<td></td>
</tr>
<tr>
<td><strong>Cash Flows:</strong> The document provides aggregate data regarding all cash inflows a company receives from both its ongoing operations and external investment sources, as well as all cash outflows that pay for business activities and investments during a given quarter (Investopedia n.d.).</td>
<td></td>
</tr>
<tr>
<td><strong>Ratios, Concepts, and Calculations:</strong> Commonly reported financial metrics extracted from the financial statements.</td>
<td></td>
</tr>
</tbody>
</table>


Websites

http://iris.thegiin.org/iris-standards#m=financialperformance
http://www.investopedia.com/terms/b/balancesheet.asp/#axzz20QSox7XB
http://www.investopedia.com/terms/c/cashflowstatement.asp#axzz20QSox7XB
http://www.investopedia.com/terms/i/incomestatement.asp#axzz20QSox7XB
http://www.investorglossary.com/annualize.htm
APPENDIX B

SMART and CREAM Indicators
The monitoring and evaluation literature identifies two standards for developing robust indicators, namely SMART indicators and CREAM indicators (Bastia, 2000; Kusek and Rist, 2004). In developing and assessing indicators from the CHMI database, we kept in mind the criteria identified by these two standards; however we did not use these criteria rigorously as we were limited by the quality of the data. The most commonly used criterion used to assess indicators was whether they were timely or time-bound. Table 3 below describes these two standards:

Table 3: General Good Practices in Developing Indicators (Busjeet n.d., 46)

<table>
<thead>
<tr>
<th>SMART Indicators</th>
<th>CREAM Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific</td>
<td>Clear - precise and unambiguous</td>
</tr>
<tr>
<td>Measurable (or Monitorable)</td>
<td>Relevant - appropriate to the subject at hand</td>
</tr>
<tr>
<td>Achievable (and/or Attributable)</td>
<td>Economic - available at a reasonable cost</td>
</tr>
<tr>
<td>Relevant (and/or Realistic)</td>
<td>Adequate - provide a sufficient basis to assess performance</td>
</tr>
<tr>
<td>Time-Bound</td>
<td>Monitorable – amenable to independent validation</td>
</tr>
</tbody>
</table>
Operation ASHA

Creating a dense network of community centers so that no patient has to walk for more than 10 minutes to obtain TB drugs and embedding these centers in accessible areas such as a grocery shop, phone booth or a temple/church – OpASHA aims to eliminate tuberculosis among disadvantaged communities.

Problem

- India has the highest incidence of tuberculosis in the world. TB rates are especially high in urban slum areas. Overcrowding increases contagion, and lack of space or access to running water can prevent adequate disposal of infected mucus that is coughed up by infected persons.
- India has 3.5 million TB patients – this represents 25% of the world’s total burden.

Types of Products / Service

- TB treatment and support lasting 6-10 months.
- MDR-TB treatment and care that lasts around 24 months for each patient.
- Over-the-counter medicines to treat side-effects of TB drugs, including nausea, vomiting, weight loss, headache, and joint pain.
- Food and nutrition supplements, contraceptives and Oral Rehydration Salt (ORS).
- OpASHA provides jobs to slum dwellers who work as counselors and providers: 60% of OpASHA’s expenses generate livelihood in the slums simultaneously with fighting TB.

Our Approach

- Treatment centres are established in close proximity to its target client base. OpASHA uses DOTS therapy to deliver TB treatments, which is an approach developed by the WHO that calls for an observer to watch the TB patient swallow their TB medication three times a week for at least two months.
- In urban areas, treatment centres are nested within existing businesses, partnering a business owner with an OpASHA counselor. The counselor distributes the TB medication, and the business owner keeps the treatment centre open while the counselor is out in the field identifying new cases of TB. The presence of the provider allows each treatment center to stay open well past normal business hours, sometime round the clock, so the patients do not have to miss work to receive their medication.
- In rural areas, counselors travel from village to village on a motorcycle/scooter, providing the DOTS therapy wherever the patient resides. These counselors also look for new cases of TB.

Our Clients

- The target population includes 5.37 million people in India and Cambodia.
- In India, OpASHA serves a population of 4.56 million in 17 cities in 6 states.
- In Cambodia, OpASHA serves a population of 1 million people in Cambodia in the 2 provinces and 4 Operational Districts.

Country of Operation: India

Year Launched: 2006

Operational / Technical Partners:

- Microsoft, MIT Poverty Action Lab, Stop TB Partnership, IFC, BigTech, Eli Lilly and Company, GivelIndia, Guide Star India, Ammado, Global Giving, Relevant Public Health Department

Health Output:

- As of 2011, OpASHA has successfully treated 18,000 patients in India and Cambodia

Number of Employees: >100

Financials

Initial Investment: N/A

Revenues:

INR 20,335,786.88 (2010-11)

Country of Registration / Legal Status: Private (not-for-profit)

Currency: INR
The Center for Health Market Innovations (CHMI) promotes policies and practices that improve privately delivered health care for the poor in low- and middle-income countries. Operated through a global network of partners since 2010, CHMI is managed by the Results for Development Institute.

http://healthmarketinnovations.org/

**Highlights**

- The default rate has seen a decrease in the areas where OpASHA is working, reaching 2.75% in 2010, as compared with the 60% default rate that was previously in place.
- The death rate for tuberculosis patients in South Delhi, where OpASHA is working, has decreased from 6% of patients in treatment to 2% of patients in treatment.
- Operation ASHA has achieved a cure rate of 91%, which exceeds the goals set by India’s Revised National Tuberculosis Control Program by six percentage points.
- Operation ASHA practices stringent cost control. The cost of treating a patient for the entire therapy of average seven months is $25.
- Each urban treatment centers is within a 1.5 km radius of the center. The centers are placed so that no patient has to walk more than 10 minutes to reach the center.
- The tuberculosis detection rate in areas served by OpASHA has increased by 95% over the four years that OpASHA has been working in these areas.

**Challenges**

- OpASHA are confronted with capacity constraints; the ability to supply the high demand for TB services is a significant challenge.

**Future Plans**

- OpASHA plans to establish 600 DOTS centers by 2012, which will serve a population of 110 million.
- The program is also set to start operation in Cambodia. The organization plans to expand outside of Phnom Phen to other parts of the country.
- The organization plans to establish 40 treatment centers in Cambodia by December 2012.
- OpASHA will install an eCompliance system in all its centers in Cambodia.
Pesinet

Pesinet uses SMS services to reduce child mortality through remote monitoring by the local doctor, thereby improving disease detection and facilitating access to medical care.

Problem

- Maternal mortality is a significant problem in many countries in Sub-Saharan Africa.
- Women lack access to health care for several reasons including lack of income, transportation, and availability of quality health facilities.
- Preventable diseases such as malaria, pneumonia and diarrheal diseases account for a significant number of child deaths in Africa. Malnutrition is responsible for 51% of deaths and is known to be a triggering or a worsening factor for other infectious diseases.
- 55% of children's mortality causes can be detected easily by periodically checking basic symptoms, in particular the evolution of their weight, which is known to be an accurate indicator of young children's health status.

Types of Products / Service

- Malnutrition tracking through SMS coupled with regular physician monitoring.
- Facilitating doctor visits as needed, coverage for examinations, and partial coverage of medications resulting from examination.

Our Approach

- Every week, Pesinet's agents visit children at home and collect simple health data (weight, fever, stools, etc.). They also provide nutritional information and illness prevention advice.
- The collected data is stored on a Java applet in the agents' mobile phones and transferred to the doctor of the partnering CSCom via mobile technologies.
- Every day, the doctor reviews the data on a web interface at the healthcare facility, identifies children at risk, and selects children needing physician visits.
- When children are called in by the doctor, families are alerted by the agent and prompted to go to the health centre. Pesinet covers the full cost of the doctor examination and half of the cost of the medication.
- Agents also organize monthly gatherings of mothers to discuss program and health issues.

Our Clients

- Clients are primarily parents of children under five in Bamako, Mali.

Country of Operation: Mali

Year Launched: 2008

Operational / Technical Partners:
- National Federation of Community Health Centers (FENASCOM)
- Ministry of Health

Health Output:
- 800 children enrolled
- 750 households benefited from prevention and health education activities during the mobile agents home-based visits
- 200 mothers participated in collective health briefing animated by the mobile agents
- Over 70% of the subscribers resort to the partnering healthcare center when their child is sick

Number of Employees: <10

Financials

Initial Investment: N/A

Revenues: N/A

Country of Registration / Legal Status: Private (not-for-profit)

Currency: USD
The Center for Health Market Innovations (CHMI) promotes policies and practices that improve privately delivered health care for the poor in low- and middle-income countries. Operated through a global network of partners since 2010, CHMI is managed by the Results for Development Institute. 

http://healthmarketinnovations.org/