



# Measuring Maternal Child Health Care in Health Results Based Financing Impact Evaluations

David Peters & Shivam Gupta  
November 3, 2010

# Background to HRBF Evaluation

- Results based financing (RBF) pilot schemes in 8 countries to link funding to improvements in maternal & child health (MDGs 4 & 5)
- Country-specific strategies with supply &/or demand-side subsidies to overcome constraints / incentivize behavior
- Rigorous, prospective impact evaluations to identify causal effects, operational feasibility, and costs of RBF schemes affecting access and quality of health care, health expenditures, and health outcomes
- Common evaluation methods that support country-based measurement systems

# Today's Objectives

- Identify common indicators for evaluation of maternal and child health care (MCH) for HRBF
  - Principles and examples
- Identify methods for assessing the quality of care by health workers

# HRBF Demand and Supply Side Incentives

## Supply Side

Monetary transfers to service providers based on:

- Number of services
- Technical quality of care
- Outcomes for patients & communities

## Demand Side

Monetary or in-kind transfers to households (often mothers) conditional on adherence to health care use or outcomes:

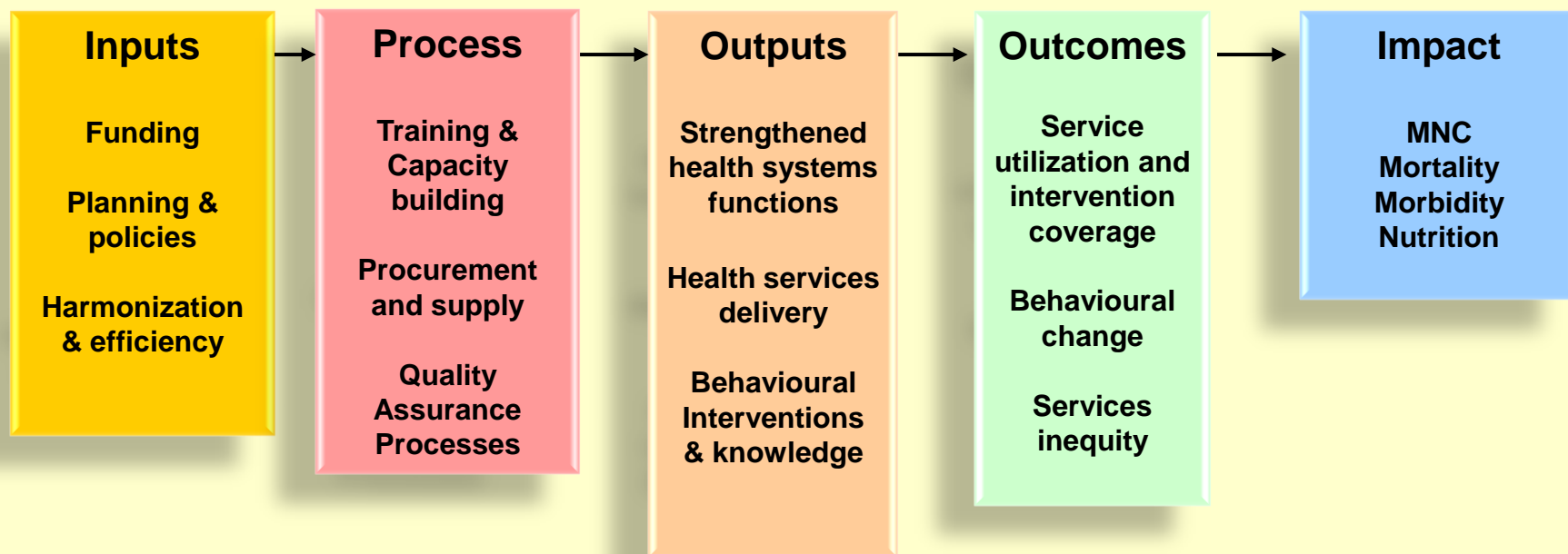
- Antenatal care
- Skilled delivery
- Childhood immunizations

Indicators and measurement for evaluation has different requirements than those used for making payments (timeliness, independence of observation, representative sampling)

# Key Elements of HRBF Evaluation Framework

1. Conceptual model specify how activities will lead to results
2. Compatible designs for evaluation
3. Standardization of common measures

# 1. The Generic Conceptual Model (IHP+)






**Demographical, Epidemiological, and Health Systems Factors**

**Political, Economic, Social, Technological, Environmental factors**

## 2. Common HRBF Impact Evaluation Design

- Country-specific policy questions & designs
- Prospective design: baseline (pre-intervention) and follow-up (post-intervention) data collection
- “Control” (comparison) areas
- Randomized allocation

				Timeline				
Program	Treatment	Design		Implement			Improve	Implement
	Control							Implement
Evaluation		Design	Baseline	Expose	Follow-up	Analyze	Feedback	

# 3. Standardization of Measures

1. Sampling methods (representativeness)
2. Indicator selection & definition (intended results)
3. Variable selection & definition (determinants and unintended consequences)
4. Data collection instruments (sources, questionnaires, training & supervision)
5. Data coding (missing values)
6. Analysis (scaling, weighting, theoretical models)



# Standardized MCH Indicators: Sources

1. Millennium Development Goals (MDGs) and targets
2. Countdown to 2015 – Maternal, Neonatal, Child Health (MNCH)
3. WHO toolkit to measure health system strengthening
4. Health Metrics Network
5. Health facility assessments (MEASURE and others)
6. MCH program documentation guidelines for the Catalytic Initiative
7. Standardized household surveys (DHS/MICS/LSMS)
8. Peer-reviewed literature on health care assessment tools

# Indicator Selection Criteria

1. Validity (measures what its supposed to measure)
2. Reliability (repeatability)
3. Relevant
  - Amenable to change as a result of intervention
  - Based on logic model (e.g. Inputs  $\Rightarrow$  Outcomes/Impact)
4. Feasible for measurement on regular basis across sites
5. Consistent with global standards
6. Limited in number

# Consensus on Standardized MCH Indicators

Consensus on Standardized Indicators

High

Low

Input

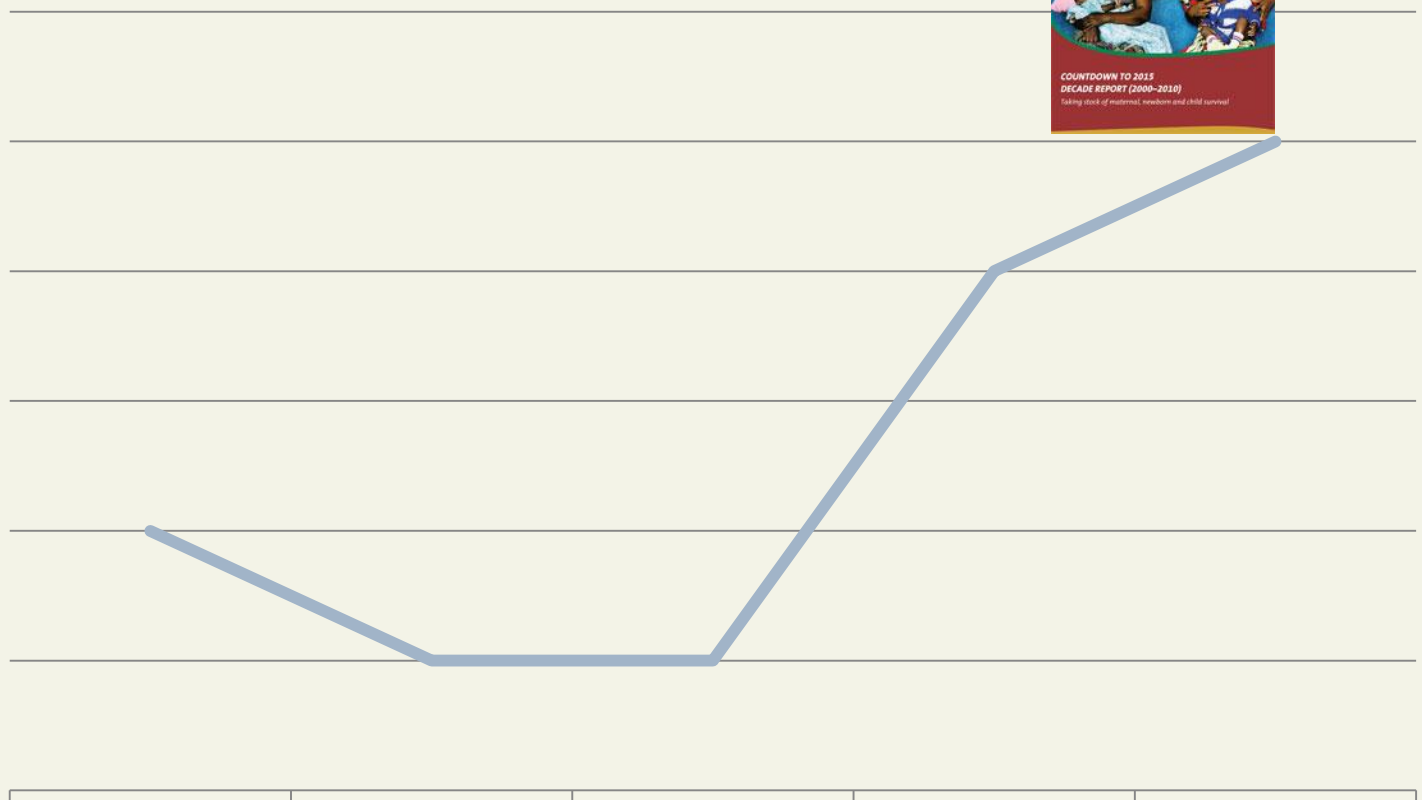
Process

Output

Outcome

Impact

Type of MCH Indicator



# Examples of Proposed Core Indicators - Inputs

<b>Indicator</b>	<b>Numerator</b>	<b>Denominator</b>	<b>Method 1</b>
Expenditure per target population (during specified time)	Total HRBF expenditure (constant purchasing power parity)	Size of target population (during specified time)	Institution based expenditure records (numerator); Census, civil registration, population based survey (denominator)
Number of long-lasting insecticide treated nets (ITNs)* purchased per target population (in specified time)	No. of ITNs purchased	Size of target population (during specified time)	Institutional financial/procurement records (numerator); Census, civil registration, population based survey (denominator)

\* Or essential drugs, vaccines

# Example of Proposed Core Indicators - Process

<b>Indicator</b>	<b>Numerator</b>	<b>Denominator</b>	<b>Method 1</b>
HW supervision status for the previous six months per trained worker*	Number of supervisory visits to health workers requiring supervision	Number of health workers requiring supervision	Institution based resource records; Facility survey

\* Can also be measured by particular health worker cadre or at health facility level

# Example of Proposed Core Indicator - Output

<b>Indicator</b>	<b>Numerator</b>	<b>Denominator</b>	<b>Method 1</b>
Proportion of standardized MCH equipment available at health facility (see list)	Proportion of standardized equipment list present and working at time of observation.	Total number of equipment on standardized equipment list	Facility survey

# List of MCH Equipment at Facility that Provides Emergency Obstetric Care

1	Sterile gloves	16	Delivery kit
2	Antiseptic liquid	17	Newborn resuscitation kit
3	Blood pressure measuring equipment	18	Timer or clock with second hand
4	Tape measure	19	Weighing scale
5	Light source (lamp or hand torch)	20	Height measure
6	Examination table or bed	21	Thermometer
7	Delivery table	22	Stethoscope
8	Intravenous sets (needles and tubing)	23	Suction/aspirating device
9	Urinary catheters	24	Stretcher
10	Fetoscope	25	Vaccine thermometer
11	Vaginal specula (small, medium, Large)	26	Cold box/vaccine carrier
12	Partograph	27	Ice packs
13	Vacuum extractor	28	Refrigerator
14	Forceps	29	Sterilization equipment (Autoclave/boiler/steamer)
15	Kit for caesarean sections	30	Puncture proof container for sharps disposal

# Core Indicators – Outcome & Impact

## Annex B Definitions of Countdown indicators



Intervention	Indicator definition	Numerator	Denominator
<b>Maternal and newborn health</b>			
Contraceptive prevalence	Percentage of women currently married or in union ages 15–49 that are using (or whose partner is using) a contraceptive method (either modern or traditional)	Number of women currently married or in union ages 15–49 years that are using (or whose partner is using ) a contraceptive method (either modern or traditional)	Total number of women ages 15–49 that are currently married or in union
Adolescent birth rate	Number of births to adolescent women per 1,000 adolescent women <sup>b</sup>	Number of live births to adolescent women ages 15–19	1,000 adolescent women ages 15–19
Unmet need for family planning	Percentage of women who are currently married or in union that have an unmet need for contraception	Number of women who are currently married or in union that are fecund and want to space their births or limit the number of children they have but that are not currently using contraception	Total number of women who are currently married or in union
Antenatal care (at least one visit)	Percentage of women attended at least once during pregnancy by skilled health personnel for reasons related to the pregnancy	Number of women attended at least once during pregnancy by skilled health personnel (doctor, nurse, midwife or auxiliary midwife) for reasons related to the pregnancy in the X years prior to the survey	Total number of women who had a live birth occurring in the same period
Antenatal care (four or more visits)	Percent of women attended at least four times during pregnancy by any provider (skilled or unskilled) for reasons related to the pregnancy	Number of women attended at least four times during pregnancy by any provider (skilled or unskilled) for reasons related to the pregnancy in the X years prior to the survey	Total number of women who had a live birth occurring in the same period
Neonatal tetanus protection	Percentage of newborns protected against tetanus	Number of mothers with a live birth in the year prior to the survey who received two doses of tetanus toxoid vaccine within the appropriate interval prior to the infant's birth	Total number of women ages 15–49 with a live birth in the year prior to the survey



# Example Core Indicators – Outcome

Indicator	Description	Numerator	Denominator	Method 1
DPT/Pentavalent immunization coverage	Percentage of children aged 12-23 months who received 3 doses of DPT vaccine/ Pentavalent	Eligible children received DPT3/Pentavalent3; according to immunization card or mother's report	Living children aged 12-23 months	Population based survey
Antibiotic treatment for pneumonia	Percentage of children aged 0-59 months with suspected pneumonia (reported cough accompanied by short, rapid breathing and/or with a fever) receiving antibiotics	Number of children aged 0-59 months with suspected pneumonia in the 2 weeks prior to the survey receiving antibiotics	Total number of children aged 0-59 months with suspected pneumonia in the 2 weeks prior to the survey	Population based survey
Skilled attendant at delivery	Percentage of births attended by skilled health personnel (country specific definition)	Eligible women delivered with a trained health care worker.	Women with a birth in previous 12m	Population based survey

# Implications for Indicator Selection

- Little consensus on standardized, cross-country Input, Process, or Output indicators
  - Highly context-specific relevance and measurement
- Few indicators have been assessed for reliability or validity
- More work on reaching consensus needed
- More testing of validity and reliability needed

# Contrast Approach with Balanced Scorecard Used in Existing RBF in Afghanistan

Management system rather than measurement system

Frontline providers, NGOs, MOPH, donors agree on:

- Purpose of Balanced Scorecard
- Domains to measure
- Unit of analysis
- Process & frequency of review/decisions
- Principles for benchmarking
- Short-listing indicators based on face validity, importance, reliability

Monitoring & Evaluation Board Final Arbitrator

Source: Hansen et al (2008). Measuring and Managing Progress in the Establishment of Basic Health Services: The Afghanistan Health Sector Balanced Scorecard. *IJHPM* 23 (2): 107-117.

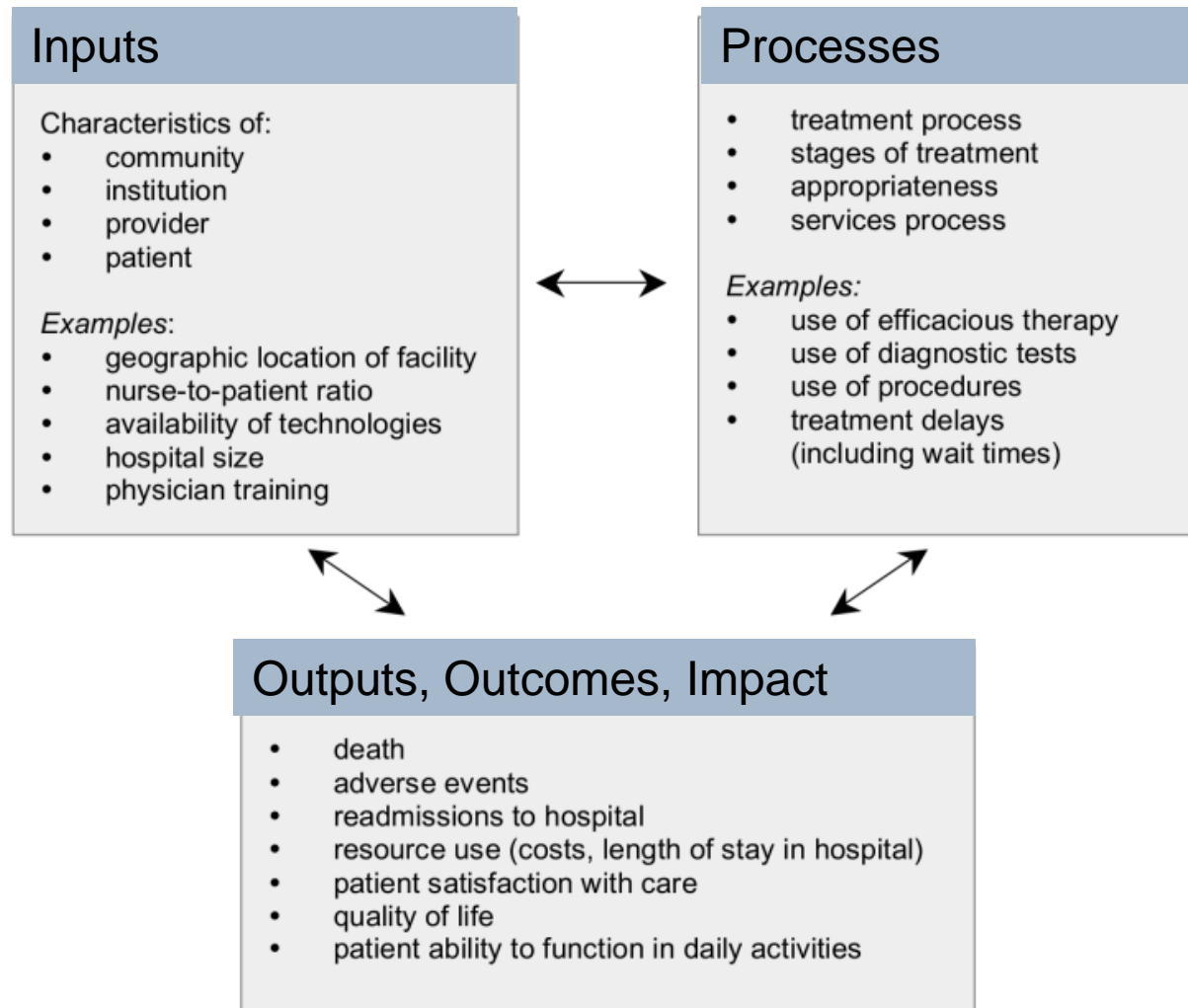
# HRBF Interventions: Why Quality of Health Worker Services is Important

- RBF Intended to directly and indirectly influence HW behavior:
  - technical quality of care
  - volume of services provided
  - coverage of services
- Not designed to measurably change health impact during evaluation period (insufficient time and size)

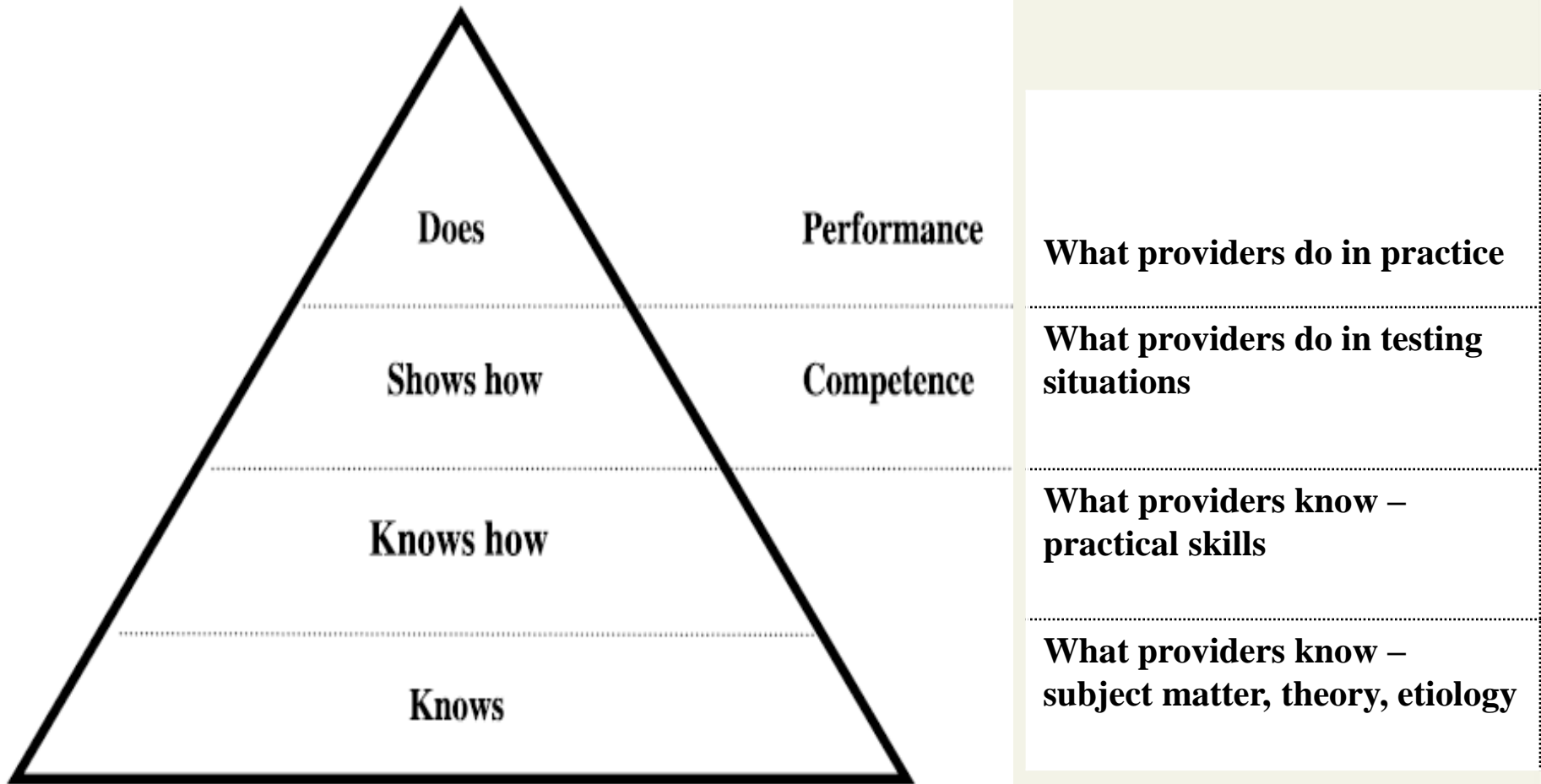
# Quality Considerations

- Quality of health care is multi-dimensional
  - Measured across multiple domains – Input to Impact
  - Multiple perspectives – patient, technical standards, managerial efficiency
- Widely accepted models
  - Health care quality (Donabedian → Performance Improvement)
  - Medical education (Osler → Competency-Performance)

# Classical Quality of Care Framework – Donabedian



# Provider Assessment – Medical Education & Practice Model



Sources: Adapted from Miller GE (1990). The assessment of clinical skills/competence/performance *Academic Medicine*; Rethans J-J et al (2002) The relationship between competence and performance: implications for assessing practice performance. *Medical Education*

# Provider Assessment – Constructs and Methods

Construct	Methods of Data Collection
<b>Performance: What providers do in practice</b> (History (Hx); Physical Exam (Px); Diagnosis (Dx); Treatment (Rx); Counseling (Cx); Professional attributes)	Simulated Patient (Mystery Patient) Patient-Provider Observation (Video or Direct) Patient Records (+ Medical Audits) Patient Exit Interviews (Reconstructed Interaction)
<b>Competency: What providers do in testing situations</b>	Patient-Provider Observation Objective Structured Clinical Exam Clinical Vignette (Role playing Hx and Cx)
<b>Practical Knowledge: What providers know about what to do (skills &amp; attributes)</b>	Clinical Vignette Clinical Case Scenario
<b>Theoretical Knowledge: What providers know about theory, etiology, subject content</b>	Written or Verbal Test



# Performance: Assessing Doctors in Routine Practice

Systematic review: 61 studies (none in LMICs)

Context: Improving doctor performance (mentoring system or problem focused)

Method	Content Validity	Reliability
Simulated Patient	High (If SP not detected: < 8%)	High (G>0.8)
Video Observation	High (if random sampled)	High (G>0.8)
Direct observation	High	G coefficient not tested (Inter-rater reliability fair - .56)
Medical Record Audit	Performance not recorded (68%) Highly variable	High (G>0.8)
Peer Assessments / Portfolio Appraisals	High	High (G>0.8)

**No assessment has been linked to patient health outcome**

# Competency: Observation of Objective Clinical Exams with Medical Trainees

Systematic Review: 55 instruments (85 studies – none in LMICs)

Context: Trainee Assessment at Training Site

- Most not feasible for large scale research
- 2 for pediatrics; 1 for obstetrics
- Few have reliability or validity measures
- None linked to patient outcomes

# Feasibility of Methods for Assessing Clinical Quality

Method	Construct	Feasibility
Simulated Patient	Performance (All: Hx, Dx, Rx & Cx + attributes)	No: Pediatric & obstetric cases not credible/ethical
Patient-Provider Observation	Performance – Competency (All)	Medium: Better when higher case load (e.g. >5 day); Common conditions
Patient Exit Interviews	Performance (All)	High: Better when higher case load (e.g. > 5 day); Limited items
Patient Records	Performance (Hx, Px, Dx, Rx, Cx)	Low: Most records in LMICs are inadequate
Objective Structured Clinical Exam	Competency (Hx, Dx, Px, Rx, Cx + attributes)	Low: Difficult for pediatric or obstetric cases; Common conditions
Clinical Vignette	Competency (Hx, Cx) Practical knowledge (Hx, Px, Dx, Rx, Cx)	Medium: All conditions; Smaller samples (need highly qualified interviewer)
Clinical Case Test	Practical knowledge (Hx,Px, Dx, Rx, Cx)	High: All conditions; Large samples
Written or Verbal Test	Theoretical + practical knowledge	High: All conditions; Large samples

# Validity and Reliability of Methods for Assessing Clinical Quality

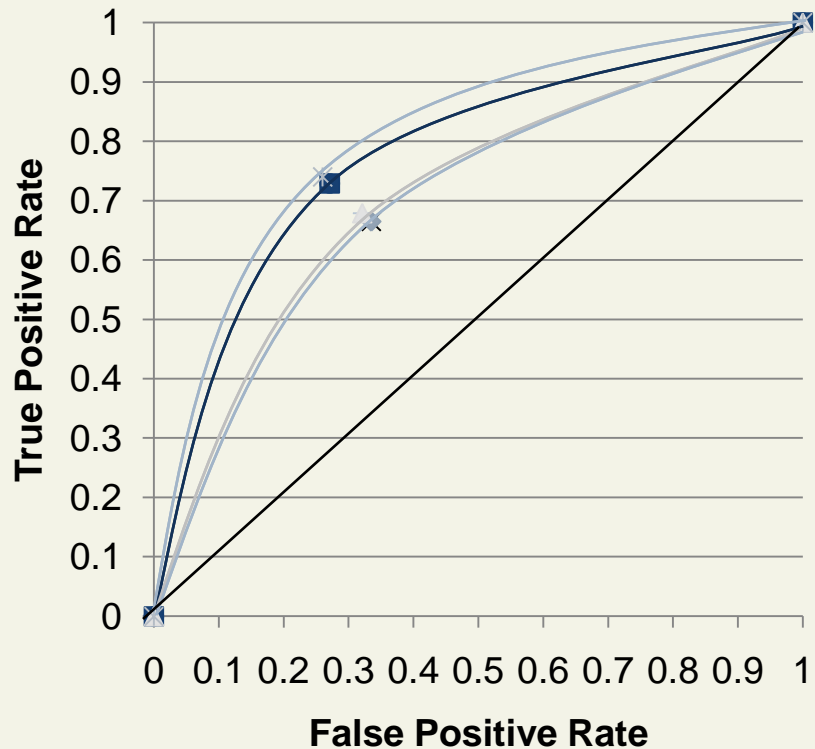
<b>Method (Construct)</b>	<b>Validity Strengths</b>	<b>Validity Weaknesses</b>	<b>Reliability</b>
Patient-Provider Observation (Performance)	Measures actual performance (usually best effort)	Observation bias (Hawthorne effect)	Potentially good
Patient Exit Interviews (Performance)	Presumed High for counseling effectiveness (& perceptions)	Poor-Fair for clinical performance (Perceptions at point of care may not sustain)	Good inter-rater
Clinical Vignette (Practical Knowledge)	Measures practical knowledge	Poor correlation with performance (does not assess behavior)	Potentially good
Clinical Case Test (Practical Knowledge)	Measures practical knowledge	Presumed poor correlation with performance (does not assess behavior)	Potentially good
Written or Verbal Test	Measures	Presumed poor correlation with performance (does not assess behavior)	Potentially good

# Comparison of Methods to Assess Quality of Pediatric Treatment of Cough, Diarrhea, and Fever

Activity	% Agreement (Range)	Kappa (Range)
<b>Comparison of Observation vs. Exit Interviews</b>		
General Assessment Tasks	43 – 97	(-0.284, 0.684)
Case Management of Cough	70 – 88	(0.114, 0.755)
Case Management of Diarrhea	49 – 97	(0.133, 0.906)
Case Management of Fever	53 – 97	(0.111, 0.796)
<b>Comparison of Observation vs. Provider Interviews</b>		
Case Management of Cough	67	0.318
Case Management of Diarrhea	62 – 82	(0.220, 0.602)
Case Management of Fever	67	0.312

Source: Franco LM, Franco C, Kumwenda N, Nkhoma W. Methods for assessing quality of provider performance in developing countries. Int J Qual Health Care 2002 Dec;14 Suppl 1:17-24.

# Receiver Operating Curves for Patient Exit Interviews vs. Clinical Observations of Pediatric Counseling: Afghanistan (2007)



	ROC	Prev (%)
× Diagnosis provided	.665	57.7
● Home advice	.729	79.8
+ Medicine adverse reactions explained	.679	14.0
- Signs for return to facility	.740	46.9

### ROC Test Accuracy Guide

>.90	Excellent
.80-.90	Good
.70-.80	Fair
.60-.70	Poor
.50	Worthless

# Assessing Clinical Quality: Recommendations for HRBF

- Use structured patient observations for assessing performance quality of common pediatric conditions (obstetric conditions under development)
- Use patient exit interviews for assessing effectiveness of counseling and other patient attributes (e.g. satisfaction, equity)
- Use vignettes (small samples) &/or case scenarios (larger samples) for assessing practical knowledge and uncommon conditions
- More research on validity & reliability

# Concluding Thoughts on a “Work in Progress”

- Consensus-building needed on selection of MCH indicators (Input, Process, Output)
  - More specific results models and hypotheses needed
- True HW performance measures are elusive
- Quality is multi-dimensional: Multiple approaches to quality measurement needed
- Investment in validity & reliability of potential indicators needed