



**USAID**  
FROM THE AMERICAN PEOPLE

TRANSLATING  
RESEARCH  
INTO ACTION



Maternal Health **Task Force**

**URC**  
UNIVERSITY  
RESEARCH CO., LLC

---

## DRAFT MEETING REPORT

# THE IMPACT OF PERFORMANCE-BASED INCENTIVES ON THE QUALITY OF MATERNAL AND NEWBORN CARE

20-21 APRIL 2016

BOSTON, MASSACHUSETTS

USAID | Translating Research into Action (TRAction) Project

Maternal Health Task Force

This report was produced for review by the United States Agency for International Development. It was prepared by URC/TRAction and the Harvard University Maternal Health Task Force.

This report is made possible by the support of the American People through the United States Agency for International Development (USAID). The contents of this report are the sole responsibility of URC and do not necessarily reflect the views of USAID or the United States Government.

---

Translating Research into Action, TRAction, is funded by United States Agency for International Development (USAID) under cooperative agreement No. GHS-A-00-09-00015-00. The project team includes prime recipient, University Research Co., LLC (URC), Harvard T.H. Chan School of Public Health, and sub-recipient research organizations.

# TABLE OF CONTENTS

## CONTENTS

Executive Summary .....	ii
Acronyms .....	iv
Session 1: Performance-based Incentive Programs: How and why do they work? When don't they work?.....	1
Session 2: Who defines quality and why does it vary: A multi-country analysis.....	2
Quality is more critical than ever .....	2
People demand quality .....	3
Not enough is known about quality, though it is critical for achieving the SDGs.....	4
Session 3: The ethics of performance-based-incentives and quality frameworks: key questions and considerations.....	4
Session 4: An Overview of World Bank PBI Programs Addressing Quality of Care and the Evidence on Impact .....	5
Measuring Quality.....	5
RBF Impact Evaluations .....	6
Session 5: The landscape of PBI programs addressing quality of care and what we can learn from the quality assessment indicators employed.....	8
Sessions 6 & 7: Country Findings.....	10
Quality of Care.....	10
Intended and Unintended Consequences.....	11
Financing .....	12
Data Collection, Analysis and Use .....	12
Political Support.....	13
Recommendations of Working Groups.....	14
A.Evidence Gaps and Implementation Research Priorities .....	14
B.Policy/Governance and Programs .....	15
C.Metrics.....	16
Session 8: Galvanizing Leadership .....	19
Appendix A: Program Agenda .....	21
Appendix B: Participant List .....	25

To learn more about TRAction, visit [www.tractionproject.org](http://www.tractionproject.org) or contact us at [tracinfo@urc-chs.com](mailto:tracinfo@urc-chs.com)

## EXECUTIVE SUMMARY

On April 20-21, 2016 USAID's Translating Research into Action (TRAction) Project and the Maternal Health Task Force (MHTF) at the Harvard T.H. Chan School of Public Health hosted a technical consultation on the impact of performance-based incentives (PBI) programs on the quality of maternal and newborn care. A total of 34 participants including researchers, programmers, policymakers, and donors gathered in Boston to review the state of the evidence and several country experiences with PBI programs. Over the course of the two days, many ideas were shared and discussion ensued on the relationship between the implementation of these schemes and the care that women and newborns are receiving. Importantly, the meeting culminated in the crafting of a set of recommendations on implementation science priorities, policy/governance and program needs, and measurement issues.

The meeting objectives were the following:

- 1) Share the state of the evidence on whether and how performance-based incentives affect quality of care, especially as applied to the field of maternal newborn health, from key low- and middle-income countries;
- 2) Understand existing indicators and methods for assessing quality (structural, adherence to clinical standards of care, patient experience) and the strengths, challenges, and considerations of each;
- 3) Consider what we know about the intended and unintended consequences of PBI programs related to:
  - a. Health workforce
  - b. Quality of patient experience
  - c. Health facility readiness;
- 4) Identify implementation research and quality measurement priorities to advance the effectiveness of PBI programs in addressing quality of care;
- 5) Consider how the state of the evidence can and should inform practice and policy moving forward for PBI interventions.

Meeting participants highlighted that increasingly, low- and middle-income countries, with support from their development partners, are investing in PBI schemes to improve health facility performance, including quality of care. With the adoption of the Sustainable Development Goals (SDGs), increasing utilization of facility-based care is a key strategy for improving maternal and newborn health outcomes and preventing death and disability. PBI is one approach being used to improve the availability and quality of facility-based care, which is in turn expected to increase service uptake. PBI also features prominently in the recently launched, multi-stakeholder Global Financing Facility, which will serve as an important financing platform in support of the SDGs and the UN's Every Woman, Every Child strategy.

While the popularity of various kinds of performance- or results-based financing schemes increases, there is much that remains unknown. As PBI programs and other strategies to increase utilization of facility-based care move forward, there is an ethical imperative to understand the quality of care being accessed by women and children in these facilities, and a programmatic imperative to understand whether the PBI investments in maternal and newborn health are resulting in improved quality of care.

The presentations and discussions were far-reaching and thought-provoking and included insights shared from country experiences in Malawi, Senegal, Tanzania, and Zimbabwe, as well as the Central American countries and southern states of Mexico that make up Mesoamerica. Additionally, participants

shared learnings from some of the earlier generation programs such as Rwanda. The meeting report that follows details the sessions and discussions that transpired. The high-level recommendations made based on discussions for advancing our understanding of and maximizing the role of PBI schemes in improving the quality of maternal and newborn health services were as follows:

#### On evidence gaps and implementation science

- 1) Design implementation research to investigate the ideal number and mix of indicators to assess facility-level quality of care in order to ensure PBI programs are incentivizing appropriately.
- 2) Given the current gap in evidence around when PBI programs are ready for scale-up and poised for sustainability, conduct implementation research on program readiness in these critical areas.
- 3) Consider the best ways to explore the effects of unintended consequences in PBI programs, and review the findings of evaluations that have looked at both positive and negative effects.
- 4) Investigate how best to leverage existing data collection systems to inform implementation of PBI programs.
- 5) Identify the most effective tools to map the system and context in which PBI programs operate.

#### On policy/governance and program needs

- 1) Frame PBI as a mechanism to implement quality maternal newborn health services across entire systems as part of universal health coverage.
- 2) Work with key stakeholders from government to own and implement PBI, making sure that there is space for local adaptation if/as needed, as well as engage civil society for accountability.
- 3) Consider mechanisms to integrate donor and national funding for PBI so that there is national-level ownership of the process (including data collection) and outcomes and a greater chance of sustainability.
- 4) Monitor carefully the implementation and experience of PBI programs (perhaps with the assistance of civil society) in order to ensure there are no unintended negative effects and, if any are discovered, that course corrections can be made efficiently.
- 5) Implement multiple ways to share learnings from PBI programs in real time especially on key issues such as provider training; recognition of common pitfalls; identification of system readiness; and pre-conditions for successful implementation of PBI.

#### On measurement issues

- 1) Establish a more robust learning community. Engage with colleagues regularly about successes and failures; nuances; learnings; and ways to catalyze progress in quality measurement.
- 2) Make manuals, checklists, and data publically available for the sake of transparency and learning; incorporate a participatory checklist revision process regularly. Work to evolve programs from mostly structural to process or outcome measures.
- 3) Identify ways to support regular review and revision of quality assessment within country programs.
- 4) Ensure issues of facility infrastructure, supply chain, and procurement processes, which affect quality but may be outside of provider/facility control, are taken into account in developing quality incentives.
- 5) Consider potential innovations in data generation and use, such as integrating data collection into existing information systems; investigating the issue of “weighting” data to clarify whether it leads to a valid quality index; and keeping the amount of data collected manageable so that it actually can be used and shared.

## ACRONYMS

ANC	<i>antenatal care</i>
ASSIST	<i>Applying Science to Strengthen and Improve Systems</i>
GFF	<i>Global Financing Facility</i>
HMIS	<i>health management information systems</i>
HNP	<i>health, nutrition, and population</i>
HSS	<i>health system strengthening</i>
IDA	<i>International Development Association</i>
IOM	<i>Institute of Medicine</i>
LMIC	<i>low- and middle-income countries</i>
M & E	<i>monitoring and evaluation</i>
MDGs	<i>Millennium Development Goals</i>
MOF	<i>Ministry of Finance</i>
MOH	<i>Ministry of Health</i>
MHTF	<i>Maternal Health Task Force</i>
NCD	<i>non-communicable disease</i>
NGO	<i>non-governmental organization</i>
NORAD	<i>Norwegian Agency for Development Cooperation</i>
P4P	<i>pay for performance</i>
PBF	<i>performance-based financing</i>
PBI	<i>performance-based incentives</i>
RBF	<i>results-based financing</i>
SDGs	<i>Sustainable Development Goals</i>
TRAction	<i>Translating Research into Action</i>
USAID	<i>United States Agency for International Development</i>
WB	<i>World Bank</i>
WHO	<i>World Health Organization</i>

## SESSION 1: PERFORMANCE-BASED INCENTIVE PROGRAMS: HOW AND WHY DO THEY WORK? WHEN DON'T THEY WORK?

In the opening session entitled Performance-based Incentive Programs: How and why do they work? When don't they work? Peter Berman explained that, in order to understand the importance of financial incentives, it is critical to think first about the broader contextual questions related to health systems. He shared a basic explanation of the concept of performance-based incentives (PBIs) and then went on to interview two experienced colleagues, Rena Eichler from Broad Branch Associates and Sebastian Bauhoff from the Center for Global Development.

Berman began by outlining the Harvard World Bank Flagship approach (Figure 1)<sup>1</sup>, explaining that the three main objectives of a health system are health outcomes, financial risk protection, and citizen satisfaction. He noted that quality, along with equity and access, is considered an intermediate outcome and is instrumental in producing the desired final outcomes. To improve the performance of the health system, governments can adjust five “control knobs”: financing, organization, payment, regulation, and behavior. Berman commented that it is important to remember that changes in the payment control knob cannot be made without identifying and mitigating the potential interactions with the other control knobs, particularly organization and regulation. Before implementing PBI schemes, countries need to conduct a comprehensive assessment of the entire health system.

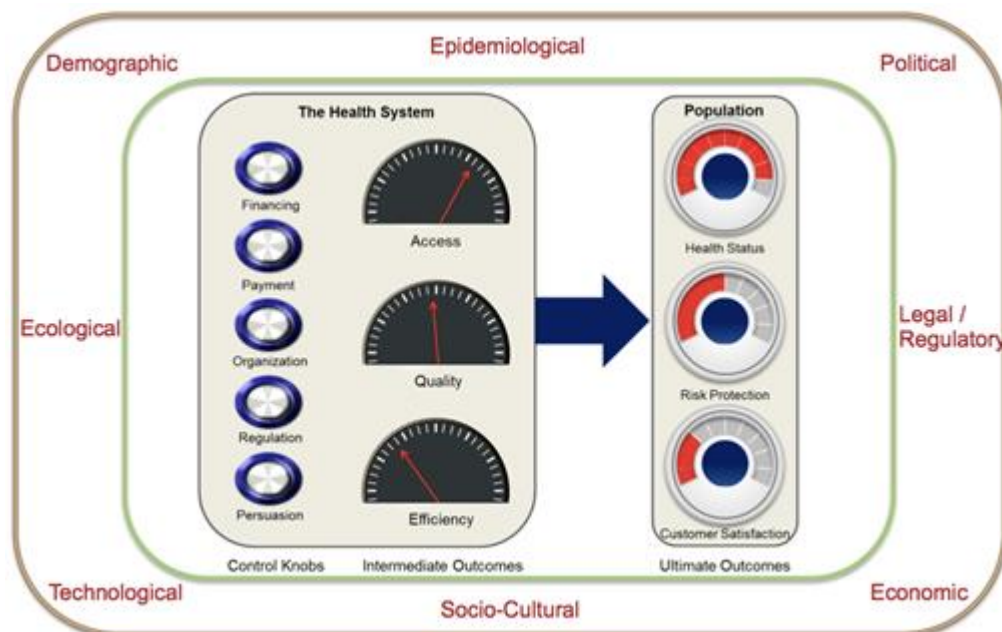


Figure 1: The Harvard World Bank Flagship Approach  
Source: Roberts et al. (2008)

Empirical evidence has shown that financial incentives are powerful motivators for health providers, Berman noted, which can influence behavior and illicit change in the quality and quantity of services provided. Incentives, however, cannot be monetized without understanding how the health system

<sup>1</sup> Roberts, M., Hsiao, W., Berman, P., and Reich, M. (2008). *Getting health reform right: a guide to improving performance and equity*. Oxford University Press.

works at the provider and managerial levels (district, state, national). When considering quality within PBI, Berman suggested that we need to consider not just *what* is rewarded, but *who* (which entities) and *how* (e.g. details of payment mechanisms and frequency). In any setting, it is important to ask what the barriers are to improved quality and who has the power to influence them.

Berman's discussion with Eichler and Bauhoff highlighted the fact that different entities within the health system can address barriers, and incentives need to be designed at all levels. PBI cannot be expected to work unless the entity being rewarded has the ability to address the underlying performance problem. As Eichler noted it is important to remember that PBI is a multi-faceted intervention which, apart from financial incentives, also includes increased supervision and improvements in infrastructure. Well-designed PBI schemes have the potential to change what providers do, as well as alter norms in the long term.

Despite the growing body of empirical evidence on PBI, Eichler and Bauhoff noted that there are still a number of important gaps that must be addressed moving forward. First, few evaluations examine the unintended consequences of PBI schemes. Although frequently there is concern that performance-based incentives are likely to result in negative unintended consequences such as providers shifting focus away from non-incentivized services, it is also plausible that these schemes could generate positive unintended consequences, particularly for services that are jointly produced. For example, rewarding one type of vaccination might lead to providers also providing other vaccinations during the same visit since the marginal cost of an additional vaccination to the provider would be small. Documenting such evidence is important. Second, the existing evidence focuses almost exclusively on the impact of PBI programs on utilization and structural quality. Apart from the immediate impact on the incentivized indicators, it is important to evaluate the extent to which PBI schemes impact patient satisfaction (or experience) and health outcomes. Third, relatively little is known about providers' motivation and the cost-effectiveness of the programs. More research should be done in these domains, particularly as PBI programs evolve and become more sophisticated.

## SESSION 2: WHO DEFINES QUALITY AND WHY DOES IT VARY: A MULTI-COUNTRY ANALYSIS

In this session, Dr. Margaret Kruk noted that improving quality of care is critical for improving health outcomes and for meeting internationally agreed upon health goals. She noted that health is a function of both service utilization and the quality of the services utilized. Quality of care can be conceptualized as technical quality (based on the effort of the clinician to provide safe and evidence-based care that adheres to accepted protocols) and interpersonal quality (including patient experiences of respect, dignity, autonomy, confidentiality, communication, choice, promptness, amenities, and social support as a part of the delivery of care). Dr. Kruk documented support for three basic premises related to quality of care: 1) that quality is more critical now than ever; 2) that people in low- and middle-income countries demand quality; and 3) that we do not know enough about quality because of the absence of quality indicators in routine monitoring efforts, including, for example, the proposed monitoring of SDG progress.

### QUALITY IS MORE CRITICAL THAN EVER

Kruk shared illustrative data from three studies that document the magnitude of quality deficiencies and their consequences:

- A global cross-sectional analysis found that even when providers had the inputs required to provide appropriate care, women were still 2-3 times more likely to die compared to the predictive model (Souza et al., 2013<sup>2</sup>);
- Roughly 20% of women experienced disrespect and abuse during childbirth in a sample of facilities in Tanzania (Kruk et al., 2014<sup>3</sup>);
- In India, patients with angina were properly-diagnosed just 42% of the time (Das et al., 2012<sup>4</sup>).

These examples point to quality shortfalls in a variety of health system settings, and to the importance of recognizing the multidimensional nature of quality. Understanding quality as a structural issue is not sufficient and quality deficiencies must also be conceptualized through provider knowledge, patient satisfaction, and clinical process.

Quality of care is also critical because of its role in determining service utilization - quality influences patient trust in the health system. Utilization depends on patients trusting that if they come to a facility, they will be provided with high quality services and will experience positive outcomes as a result of the treatment provided. In Liberia, it was found that only 51% think that they can receive the help they need from the local health system (Svoronos, Macauley and Kruk, 2014<sup>5</sup>). Distrust in the health system is caused by many factors; satisfaction with individuals' last visit in that health facility is an important predictor, which further supports the inclusion of patient experience as an essential component of quality.

## **PEOPLE DEMAND QUALITY**

Not only is quality a critical driver of increases in service utilization, but Kruk's investigations using discrete choice experiments have found that patients demand quality and choose where they will seek care based on perceptions of quality. Studies in Tanzania and Ethiopia demonstrated that the availability of drugs and medical equipment (structural quality) and good provider interpersonal skills (respectful care) were more important to patients than distance to the health facility, cost of services, or availability of transport.

This research demonstrates that women often place a higher priority on quality than on other issues when making care-seeking decisions. For example, women may bypass a closer clinic in favor of one further away if they perceive the quality to be higher at the more distant facility. Women were more likely to deliver in health centers if the centers have been upgraded recently or have the capacity to provide three or more signal functions. Importantly, evidence confirms that women care about the quality of care provided at the facility and, in this way, trust may be the most important "outcome" of the health system.

---

<sup>2</sup> [http://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(13\)60686-8/fulltext](http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(13)60686-8/fulltext)

<sup>3</sup> <http://heapol.oxfordjournals.org/content/early/2014/09/21/heapol.czu079.full>

<sup>4</sup> <http://content.healthaffairs.org/content/31/12/2774.long>

<sup>5</sup> Svoronos T, Macauley RJ, Kruk ME. Can the health system deliver? Determinants of rural Liberians' confidence in health care. Health Policy Plan. Jul 27 2014.



### **NOT ENOUGH IS KNOWN ABOUT QUALITY, THOUGH IT IS CRITICAL FOR ACHIEVING THE SDGs**

Kruk noted that while the Millennium Development Goals led to a proliferation of research on utilization of health care, there has been relatively little research on the quality of care that individuals are receiving. While data indicate that more women are delivering in facilities now than 15-25 years ago, quality of care as measured by compliance with clinical protocols and routine signal functions is low. In spite of strong concern with poor service quality, we do not know enough about it, such as the elements of quality that are most critical, how quality varies across health systems, minimum thresholds of quality, and the best measure of quality.

There is an underutilized potential for existing data to inform our understanding of quality. For example, ongoing research by Kruk, et al. find data from five African countries indicate the technical quality of obstetric care is consistently lower in primary-level facilities compared to secondary level facilities. Additionally, the quality of sick child care appears to be lower than the quality of antenatal care in these countries. It is therefore important to look at specific service dimensions in order to understand the true variance in quality across and between countries. Leveraging and improving existing data to more routinely and effectively assess quality will be important in the Sustainable Development Goal (SDG) era.

The global shift from the MDGs to the SDGs includes a more comprehensive set of health targets, which require increased complexity of health services and place new quantity and quality demands on health systems and providers. Importantly, quality indicators are not included among the health-related SDG indicators, which address only utilization and health outcomes. Kruk advocated for creativity in examining and understanding quality of care in order to change health systems and ultimately the kind of care that women and newborns (as well as others) receive.

### **SESSION 3: THE ETHICS OF PERFORMANCE-BASED-INCENTIVES AND QUALITY FRAMEWORKS: KEY QUESTIONS AND CONSIDERATIONS**

Nir Eyal from the Harvard T.H. Chan School of Public Health provided an interesting foundation for the discussion of the impact of PBIs on the quality of maternal newborn care in a session he led entitled *The ethics of performance-based-incentives and quality frameworks: key questions and considerations*.

When designing and implementing a program that utilizes PBI, Eyal noted that a key question to consider is **whether the intervention improves the health system as a whole and whether it improves the quality of care overall**. He went on to use WHO's working definition for quality that describes six dimensions that are useful as a framework to analyze ethical considerations. These dimensions are: Is the PBI intervention effective, efficient, accessible, acceptable/patient centered, equitable and safe?

For example, he noted that PBI can cause misreporting (USAID 2010) and other gaming (Van Herck, et al., 2010) which may not be an effective way to improve quality of care. Concerns are voiced that demand-side interventions may introduce perverse incentives, for example, encouraging women to have additional children because of the cash benefits in a poorly designed conditional cash transfer program, or there may be payment schemes to providers that result in excessive provision of unnecessary or potentially harmful services (USAID 2010). This might not lead to safe maternal or child health. PBI should be used as a tool that enables better management practices by setting benchmarks, as well as increasing accountability and transparency in the healthcare system.

Beyond quality of care there are also broader ethical questions that arise for providers as well as clients. For example, workers may be unfairly penalized for not achieving outcomes due to barriers outside their control and responsibility (e.g. stock-outs) and clients may be pressured to use a form of contraception that is unacceptable to them instead of being given a selection from which they can make an informed choice.

Eyal asked provocative questions, for example, in the long term, do PBI schemes introduce unhealthy trends like commodification and micro-management or undermine trust in healthcare providers and promote excessive competition amongst workers? Could they introduce and enable corruption (especially with financial incentives)? Eyal led a brief discussion that surfaced the need for more robust evidence in this area.

Finally, he noted that the benefits of implementing PBI programs may outweigh small negative effects they may have or the negative effects may be relevant only in some forms of PBI. Thoughtful design and implementation of programs that utilize PBI schemes may enable the avoidance of unethical practice thereby contributing to improved quality of care and better functioning of the entire healthcare system.

Points from Eyal's presentation emerged in discussion throughout the remainder of the meeting as additional comments about contexts, drivers, and outcomes ensued.

## **SESSION 4: AN OVERVIEW OF WORLD BANK PBI PROGRAMS ADDRESSING QUALITY OF CARE AND THE EVIDENCE ON IMPACT**

Dinesh Nair and Damien de Walque noted that the World Bank's RBF portfolio spans 36 programs in 30 countries and \$396 million in committed investments since 2008 from the Health Research Innovation Trust Fund. These RBF investments are linked with an additional \$2.2 billion in concessional financing from the International Development Association (IDA), and aim to leverage stronger results from the basic health, nutrition, and population (HNP) services supported by IDA credits. All of these programs involve facility-level (supply-side) incentives, and some are supplemented by demand-side interventions. The majority of these – 80 percent – are in Africa. Most World Bank-supported RBF programs pay for both quantity (utilization) and quality, though some countries' programs focus on paying for quality only, and not for utilization (e.g. Liberia and Kyrgyzstan).

### **MEASURING QUALITY**

Quality measurement in RBF programs is typically fulfilled quarterly, and counter verification of checklist scores takes place either quarterly or bi-annually. Methods used to measure quality in the Bank's RBF programs include a majority focus on structural quality, though these instruments are continually evolving to include more process indicators and to more accurately measure and incentivize quality improvements overall.

Citing Liberia as an example, Nair shared that in some programs, quality is assessed through distinct checklists – one for management and structural quality, which includes use of patient chart reviews to assess clinical management of key maternal and newborn health (MNH) conditions (e.g. obstructed labor) – and another for process of care quality. Some country program checklists include patient satisfaction as a component of quality, including Nigeria and Kyrgyzstan. Quality measurement tools are not static and are continuously updated within country programs.

Several programs have measured improvements in both structural and process quality indicators following the introduction of RBF, including Benin, Burundi, and Kyrgyzstan. Lessons learned highlighted by particular country experiences include:

- **Systems-level factors that are critical for quality improvement must be addressed to support facility-level quality improvement efforts:** Results from Burundi showed all regions starting at different levels, but over time quality scores converged around 80%. This quality plateau at 80% suggests facilities face barriers to improve beyond that plateau that should be explored. Quality plateaus may result because certain barriers to improving quality are under district-level control (e.g. staffing, procurement, record systems) and cannot be addressed at the facility level. This situation should produce a bottom-up pressure on district-level management to address the systemic issues influencing quality improvement so that facilities can achieve higher quality scores.
- **As facilities achieve incentivized quality improvements, indicator weights should be adjusted to lower the importance of easier-to-reach indicators, and to increase the importance of harder-to-reach indicators:** An initial focus of checklists and indicator weights on ‘lower hanging fruit’ in RBF programs, followed by shifts to progressively incentivize more challenging quality improvements, allows incentives to keep pace with increasing facility capacity. It can be challenging to continuously adjust indicator weights as facility performance improves, but these adjustments are critical so that additional improvements are always incentivized until the desired quality level is reached.
- **Financial autonomy at the facility level is critical to enable leveraging of bonus payments to improve facility performance:** Results from Benin, which demonstrated almost no difference in quality improvement between RBF and control facilities, indicate that a lack of autonomy over funds at the facility level can impede quality improvement by blocking the movement of resources to address actual service provision issues. Clearly, as World Bank colleagues explained, there is a need for autonomy and improved supervision in addition to money.

## **RBF IMPACT EVALUATIONS**

Dr. de Walque noted that the World Bank is conducting impact evaluations of all of its RBF programs using mixed methods within rigorous evaluation designs, including randomized controlled trials or difference-in-difference models. Objectives of the RBF impact evaluations are to improve MNH in LMICs by generating and applying evidence, improving methods and measurement, building research capacity in countries, and communicating the knowledge and evidence generated.

The Bank is currently in the “third generation” of impact evaluations, with the earlier generations having investigated:

1. Does RBF work?
2. Does some variation in RBF work better than another?

3. How does RBF compare to other health system interventions, and how does RBF interact with other health system interventions?

This third generation of impact evaluation has focused on measurement of quality of care. The fourth generation of impact evaluation is yet to be defined, but may address the increasingly pressing issues of sustainability and cost effectiveness, along with comparison of demand-side vs supply-side interventions and their impact on a key set of indicators.

Recent RBF impact evaluations have found that RBF does improve quality in a number of MNH indicators. For example, in Rwanda, in addition to increasing coverage of MNH services, RBF resulted in an increase in prenatal care quality (based on prenatal effort score). In Zimbabwe, RBF was associated with an increase in the number of facility deliveries, as well as the quality of antenatal care, with increased inclusion of critical components of ANC visits such as urine sample collection and tetanus vaccine administration.

In response to questions from the meeting participants, colleagues from the Bank also made clear that most of the evaluation focus right now is on short-term gain such as rise in utilization of services and quality of care, but that there is very little information currently on outcomes such as morbidity and mortality. Drs. Nair and de Walque explained that three reasons account for this: 1) insufficient sample size; 2) timeframes that are too short; and 3) ethical considerations regarding how long a control group can be maintained if and when positive results surface.

**Using impact evaluations to develop innovations in quality measurement:** The World Bank also aims to use impact evaluations to develop new quality measurement tools that can be used in routine RBF quality measurement, de Walque explained. In Kyrgyzstan, the quality checklist is based on peer-to-peer assessment, with district health teams supervising the peer review process. The Kyrgyzstan checklists employs six different quality measurement methods, with innovative techniques like simulations of key clinical MNH care processes (e.g. postpartum hemorrhage and newborn resuscitation) to assess process quality, as well as phone interviews to assess patient satisfaction. The peer-review process in Kyrgyzstan has demonstrated positive peer effects, where providers are encouraging each other to perform better.

De Walque also shared that in Burkina Faso a “lab-in-the-field” experiment developed video vignettes to test provider knowledge. He explained that traditional vignettes describe symptoms on paper for providers to diagnose the problem. The video vignettes portray actors posing as patients representing a variety of economic backgrounds and with different health issues. The advantage of the video medium is the ability to test if the health professional’s response differs based on the perceived wealth of a patient. In Burkina Faso, there was an additional bonus for serving the poor, so the vignettes were meant to test whether providers responded differently to poor versus non-poor patients. The video vignettes found that, in facilities where the accuracy of diagnosis is not incentivized, the quality of the answer is not as good for the poor actress. However, if the correct response is incentivized, the quality of the answer is the same for both groups.

The Bank aims to test these video vignettes in comparison to paper-based vignettes. These methods measure what providers know about how to treat patients but not what providers actually do in practice. Direct clinical observations are the only option for measuring the know-do gap, though there is a risk of “observation effect.” All methods have limitations. Possible solutions may be to “plant” an observer for two days to see if the provider becomes comfortable with the observer over time or to “plant” a patient.

**Qualitative research in impact evaluations:** Qualitative learning often supports these impact evaluations, shedding light on how PBI is understood by actors within the health system, the process of program design and indicator selection, and whether the program is functioning smoothly from the perspectives of patients, providers, and supervisors.

Effects on worker motivation are mixed. Some providers view the scenario as either supportive or controlling. Some perceive the reward as too big or too small or fair or unfair. While facility autonomy may be perceived as improved in general, sometimes staff members feel the facility director does not allow staff to make autonomous decisions with the reward. The issue of an increased workload is noted by many, according to Nair, and while this may be viewed as bringing in more incentive rewards, it may also lead to provider burnout. Dr. Nair explained that they were learning in rural areas that implementing a program can result in increased teamwork, as compared to peri-urban facilities where that did not seem to be true. He suggested that providers interacting with community members also builds confidence among the community.

Additionally, supervision is noted as appreciated but, at times, potentially too controlling. With complex rules accompanying most RBF programs, they are sometimes considered too complex. Staff frequently do not understand how rewards are calculated and that lack of understanding may breed mistrust, so in its programs the World Bank aims to clarify rules and communicate them to all stakeholders.

Dr. Nair noted that for the Bank, learning what works and why is an ongoing process, and measuring quality is more complicated and complex than previously understood. Importantly, he cited the role of training and retraining in assuring a smooth process as programs are implemented. He also noted the need to take a step back and consider carefully the broader health system in which facilities operate as there may be critical pressures or issues not obvious at first glance. Noting the “black box” of human behavior, he suggested that much more exploration needs to happen in order to truly understand provider and client motivation.

## **SESSION 5: THE LANDSCAPE OF PBI PROGRAMS ADDRESSING QUALITY OF CARE AND WHAT WE CAN LEARN FROM THE QUALITY ASSESSMENT INDICATORS EMPLOYED**

Jessica Gergen presented ThinkWell’s global landscape analysis of PBI programs with a quality of care component. ThinkWell analyzed all quality of care indicators from 32 programs implemented in 29 low-

and middle-income countries. Together, program manuals and checklists from these 32 PBI programs contain more than 8,500 individual quality indicators. ThinkWell compiled the indicators into a database and categorized them by indicator typology, Institute of Medicine (IOM) quality dimensions, and means of quality assessment. ThinkWell also developed their own sub-typology to further disentangle the range of indicators within each type.

Some key findings shared in this session included the following:

- 80% of the indicators analyzed were structural, 20% were process, and less than 1% were outcome.
- The majority of indicators were related to infrastructure and equipment (48%), facility management (25%) and the presence of pharmaceuticals (8%). Together these account for 80% of all indicators.
- Indicators looking at the competence and effort of clinicians accounted for 18% of all quality of care indicators among the 68 quality checklists analyzed. These were tested through register and chart review (14%), questioning the clinician (<1%), exit interviews with patients (<1%), and direct clinical observation (3%).
- Checklists tended to more heavily weight facility management, maternal, newborn, and child care, and facility equipment and infrastructure services or domains.
- ThinkWell found very subtle, almost insubstantial, changes in the composition of checklists over time in the same PBI programs. For example, a program may have changed the number of indicators by either condensing or separating groups of indicators that relate to a single clinical process, but the proportion of structural/process/outcome indicators remained the same.
- The newest programs (DRC and Malawi) have increased the proportion of direct observation or patient-focused indicators.

Gergen highlighted some challenges in completing this analysis. One major limitation was access to program checklists and manuals, most of which are not publicly available. In addition, the study team individually coded all 8,500 indicators, which was time- and resource-intensive. Determining standardized categories for the indicators required determining an appropriate QoC framework, and the team quickly realized that categorizing certain indicators required a high level of subjectivity. For example, structural indicators are fairly straightforward while the definition of process indicators is less clear and more subjective.

The ensuing discussion noted that participants were interested to learn if checklists were all the same length, or if some were shorter than others. In addition, they asked if short indicators could replace longer ones. These questions speak to the issue of easing the burden of checklist completion for verifiers and counter-verifiers. One participant suggested analyzing how programs are weighting indicators by service or indicator type. Little is understood about how countries determine weightings.

Another participant asked if the manuals contain guidance on how to provide feedback to the providers and facilities. Through their case studies, ThinkWell has found that, in general, providers do not understand how their behavior affects their scores, nor how their scores affect their payment.

There was also a discussion about the fact that many programs only measure services that are directly incentivized and, therefore, it is not possible to see unintended effects on other service areas. Checklists

do not include data collection on indicators that are not directly incentivized. For example, a program that incentivizes maternal newborn health care does not likely measure quality indicators around tuberculosis or services related to non-communicable diseases (NCDs)

Finally, one participant asked if programs set targets for these indicators at the provider or facility levels since, in general in the PBI community, there is not a focus on target setting as much as within the pay for performance community.

## SESSIONS 6 & 7: COUNTRY FINDINGS

Learnings from a number of countries' experiences with PBI programs were shared during this technical consultation. The countries discussed included Malawi, Senegal, Tanzania, and Zimbabwe. Additionally, learnings from central American countries and southern states in Mexico were shared as part of the Salud Mesoamérica Initiative. Finally, earlier work in PBI was also represented by those familiar with earlier generation experiences, such as in Rwanda. World Bank colleagues also shared examples, as relevant, from countries such as Nigeria, the Central African Republic, and Cameroon. The country discussions were extremely rich and, in order to share important issues that emerged in multiple countries, the following section is organized by these themes: intended and unintended consequences; financing; data collection, analysis and use; quality of care; and political support.

### QUALITY OF CARE

Each of the countries presented critical information on one of the primary goals of their PBI programs: improving the quality of the healthcare services provided. They also noted the complexity of trying to measure actual quality of care.

In terms of **the impact of PBI schemes on quality of MNH**, the results are mixed. In Malawi, the RBF4MNH Initiative substantially reduced time to care for women with complications and increased the capacity of the health system to handle complications at the Basic Emergency Obstetric Care (BEmOC) level, both of which generate efficiencies with important implications for morbidity & disease burden RBF4MNH. Quality improvements attributable to the intervention were also observed in intrapartum HIV status assessment, availability of PMTCT medications, and infection prevention equipment, supplies, and processes, but not in other key quality areas like obstetric care management or client satisfaction with care. In Zimbabwe RBF was found to improve the quality of ANC care in comparison to control facilities, including the likelihood of having their urine tested and receiving a tetanus shot.

In terms of **patient perspectives of quality of interpersonal care**, in the Tanzania pilot, PBI was found to increase staff kindness during delivery and overall patient satisfaction with care. In the Malawi example on the other hand, no effect of the PBI intervention was found on patient satisfaction.

It is important for sustainability to understand the expected and unexpected ways incentives influence provider behavior, and why. For example, in Senegal, the motivation to use bleach as a disinfectant seemed to be tied to the idea that the PBF team might stop by at any moment, rather than being recognized as an important mechanism for infection prevention. While the framing of the motivation to use the bleach was skewed from the original program intent, for the sake of implementing this particular program, bleach was actually used, which was a desired practice. More broadly, however, this raises the question of how this particular type of motivation might affect the sustainability of the practice of using bleach for infection prevention.

World Bank colleagues are currently analyzing results on **process indicators as a proxy for quality of care** in Argentina, Cameroon, the Central African Republic, the Democratic Republic of Congo, Nigeria, Zambia and Zimbabwe to try to ascertain issues around provider competence and sustainability when procedures are performed according to standards. In Zambia, for example, the impact evaluation was able to document an increase in the number of women whose abdomens were palpated in an antenatal care (ANC) visit, as well as the uterine height measured. Whether either is correlated with better health outcomes for these particular women and babies, however, remains unknown, highlighting the need to assess outcome indicators of quality in addition to structural and process indicators.

From the provider side, PBI programs may be correlated with **important positive changes in workplace culture**. In Senegal, for example, providers noted that in locations with PBI schemes, there is a better work environment, not only in facility conditions, but also in terms of the workload. They also noted better communication among staff. The staff have been challenged to make changes relying on their own creativity and have risen to the occasion including rewarding the best community health workers by redistributing the rewards and considering and forecasting demand for commodities in order to avoid stockouts. They note that they are more engaged in some of the educational opportunities, as well. The impact evaluation of Malawi's RBF4MNH program also found health workers' satisfaction with their work environment and their overall job satisfaction increased significantly. Providers also reported adherence to clinical protocols requires more time and resources per patient, getting at the additional costs, including time costs, of providing better quality care.

For most of the country experiences shared (Senegal, Tanzania, Zimbabwe, Malawi) there was concern expressed about a **plateau effect** when they reached a certain score on their quality checklists and overall ratings. Often they cited fairly quick and important improvements upfront, such as with structural indicators, but when the more challenging behavioral changes were introduced, including training and capacity building, there would often be a lag before any improvements were seen. There was much discussion about the "know-can-do" gap and some ideas about how to address it. In Pwani, Tanzania, for example, the district managers decided to redeploy more efficient workers to locations where providers were struggling to meet targets. Process indicators then became more important and these are, by definition, harder to measure.

## **INTENDED AND UNINTENDED CONSEQUENCES**

Each country program shared had a specific set of objectives or goals generally including the aim to improve both utilization and quality of essential health services. Most countries' PBF programs started with a focus on **structural improvements** (as in Zimbabwe, Senegal, Tanzania, and Malawi) and the consequences were fairly easily identified and measured. For example, the Tanzania country program was able to document a much improved commodity flow as a result of the PBI program. In the Malawi and Senegal programs, a key learning was how important the structural improvements were at the beginning in order to ensure that each facility would have the infrastructure, supplies, and ultimately capacity to provide high quality services. As an increased number of facilities acquired the necessary infrastructure, the focus then shifted to using the resources appropriately to improve service quality and health outcomes. The Malawi program noted that the intended structural improvements were significant and easy to identify, but that actual improvements in clinical quality have been much more difficult to achieve.



Some countries also created very explicit goals around **data collection and management** (see more in section below on this theme, as well). The Mesoamerican countries intentionally built data collection systems to track their progress, as well as to check on any unintended effects. Interestingly, the “homegrown” nature of the data collection tools has led to countries being highly invested in this aspect of their PBI programs and they report being very interested in continuing to track their progress beyond the lifetime of the actual project. One country director explained that their reputation was at stake. That they would consider their reputations as being a motivating force was not actually considered at the outset, but has emerged as a very positive and unintended outcome. Countries and states that are doing well are invested in maintaining their reputation as leaders.

The consideration of **which levels of care to reward and the managing of relationships across and between levels and institutions** has also had some interesting consequences. The Mesoamerican programs have learned, for example, that a focus on rewarding quality not only at the clinical level, but also at higher administrative levels has proved important for maintaining effort. The administrative staff members have been critical to support the success of the programs clinically. Additionally, some World Bank colleagues shared that they have learned from earlier generation projects that an unintended and very positive consequence has been improved relations between staff from the Ministry of Health and the Ministry of Finance because they often need to work closely together to implement PBI schemes.

## **FINANCING**

Several country programs shared the fact that **domestic financing** was important for stakeholders to “buy in” to the concept of PBI scheme implementation. Within the Salud Mesoamérica Initiative, this theme emerged multiple times and was cited as evidence of strong political will critical to carry out the programs (see also section below on political support). In Zimbabwe, the PBI program is also supported by Government resources in addition to donor contributions and World Bank funding, resulting from the increased government leadership and ownership for PBI that have developed over time, also reflected in the National Results-Based Management Strategy and the Ministry of Finance’s Results-Based Budgeting approach. Meeting participants also stressed that basic financing from the government (or other source) cannot and does not make up for whatever resources are needed simply to maintain health systems. PBI schemes cannot be expected to make up for what should be basic support provided, an issue that resonated again in the discussions around quality. In Malawi, for example, there is a focus now on domestic support in terms of leveraging additional government funding to scale up the programs to date. This effort appears to be in parallel with strengthened governance and coordination of the role of government at the national level.

## **DATA COLLECTION, ANALYSIS AND USE**

Each country program had specific and important ideas about data and the one universal thread was the need for programs **not to get so invested in collecting data that too much information was collected** and the result simply overwhelmed both the institutions and the programs involved. One participant noted that when this happens, the only recourse is to agree to eliminate large portions of the data gathered and to try to take a fresh look at what data are important and can be used to inform the program. As noted above, creating the tools to gather, analyze and use the data may be best if designed by those who will actually do the gathering and analysis, as the Salud Mesoamérica Initiative has learned.

Additional insights included the **role of data collection and verification** as an important step to making sure that data are valid and used to inform the payment schemes and these processes look different in different places. In Senegal, for example, data are collected at the lowest level of care (health post or center), imported into a report that includes their quality checklist, and then submitted to the district. The district then compiles all of the reports and submits to a regional management center which manages a data verification process before payments are issued. The actual verification process is taken very seriously and has two components: 1) staff from the regional management center examines the quantity-based performance indicators and the scores on quality checklists from a sample of the facilities; and 2) local non-governmental organizations then take these data, checking with register books with the local facilities selected and tracing a subset of facility users to make sure that they did, in fact, seek services during the time and at the location specified.

In Tanzania, although the process used has been slightly different, important insights have also been gleaned about the role and **toll of the verification component**, including that health workers ended up spending 17% of their time monthly on data generation and verification and this was a considerable threat to the amount of time that they have available to actually provide care to patients.

World Bank colleagues have been trying to determine who is best placed to take on the verification process and have learned that district health teams may be “too close for comfort” to the staff and data being evaluated and it may not be fair to either party to charge them with the responsibility for verification.

## **POLITICAL SUPPORT**

Finally, countries cited political support as being absolutely **crucial to programs’ success** and shared a number of different ways this was achieved.

In Malawi, for example, a “**bottom-up” approach** was used to introduce the PBI programs; district level stakeholders played critical roles introducing results-based schemes at both the community level and with the Ministries. They became the voice representing the experience of PBI schemes to the central authorities and found that national-level support followed when they had demonstrated positive results at the lower levels.

In Senegal, where the program was **designed locally** (e.g. not brought in by a donor) following an inspiring trip by Senegal’s Minister of Health to learn from Rwanda’s success, much credit is given to those who instigated the idea. Their approach was highly transparent, involving all stakeholders across levels and areas of expertise from the very beginning. In every design and review meeting, all stakeholders were welcome and the communities were invited. As results began to come in, a similar approach was taken and the political support generated was significant.

In the Salud Mesoamérica Initiative, there have been fundamental lessons learned about ownership and **champions** who have emerged for the programs. These champions have often been able to get the attention of key stakeholders locally, as well as have the ear of politicians or high level administrators who are able to leverage additional support needed. For each champion, there has also been some outreach to key individuals for domestic financing so that governments are not only perceived as supportive, but actually are in the role of finding and earmarking funds for PBI programs.

## RECOMMENDATIONS OF WORKING GROUPS

### A. EVIDENCE GAPS AND IMPLEMENTATION RESEARCH PRIORITIES

The small group working on evidence gaps and implementation research discussed a number of key areas in the application of PBI programs for improved quality within maternal newborn care that deserve further investigation. The following summarizes the key themes that form the basis of an implementation research agenda for PBI programs that address quality of care.

It is clear from the literature and from country experiences reviewed that the number and mix of quality assessment indicators used across PBI programs varies substantially. There is little evidence to inform the ideal number and mix of indicators required to assess particular quality aspects or to comprehensively assess quality of care provided in the context of a particular facility. Which aspects of quality can and should be incentivized and monitored as part of PBI schemes? What is the right mix of indicators and how should that mix shift over time? Is it possible for a shorter list of key indicators to accurately measure quality of care, allowing for the time and resource intensity of quality assessment and verification to be reduced?

Additionally, the group highlighted the fact that research on how and why PBI programs that incentivize care quality work is critical to understand, as well as why they work. They noted that a clear sense of what drives effectiveness of PBI program implementation, scale up and sustainability, will be important to clarify. Equally important is the assessment of the unintended consequences of PBI programs. The group consensus was that more could be done to document and understand the positive and negative consequences of PBI design choices—and there is an ethical imperative to do so.

Finally, working on the data systems and the tools to inform these systems was cited as key. The group suggested the need to leverage existing data systems, including HMIS, DSS, and SPA, to inform implementation and impact. They also wanted to know what tools could help map the systems and context in which PBI programs operate noting that understanding context is essential to the design and successful implementation of PBI programs that aim to improve quality of care.

The group agreed more should be done to document and disseminate important implementation lessons learned. They felt that it was critical to determine how to incentivize and create space for learning agendas and implementation research activities within the context of PBI program implementation including learning from successes and failures. The Salud Mesoamérica Initiative presented a promising model and their example led to discussion of establishing a repository or other collaborative space for sharing lessons learned. They also noted that considering whether PBI programs are more effective or better suited to create lasting change--as compared to quality improvement efforts is important. The development of a refined implementation research agenda with broader stakeholder buy-in was also cited as a priority.

With these discussion points in mind, the following are the group's five recommendations:

#### ***Recommendations***

- 1) Design implementation research that will investigate the ideal number and mix of indicators to determine quality of care at a facility level in order to ensure that PBI programs are incentivizing the right variables.

- 2) Given the current gap in evidence around when PBI programs are ready for scale-up and poised for sustainability, craft implementation research on program readiness on these two fronts.
- 3) Consider the best ways to explore the effects of unintended consequences in PBI programs, and review the findings of evaluations that have looked at both positive and negative effects.
- 4) Investigate how best to leverage existing data collection systems to inform implementation of PBI programs.
- 5) Identify the most effective tools to map the system and context in which PBI programs operate.

## **B. POLICY/GOVERNANCE AND PROGRAMS**

Five areas were discussed in relation to the policy and governance considerations for PBI programs and their role in quality of maternal and newborn care: scope; stakeholders; sustainability and financing; ethics; and learning and sharing. The group decided to adapt their recommendations to address what specifically the key policy/governance and program gaps are when considering how performance-based incentives fit into the strategies to achieve universal health care.

### ***Scope***

The group agreed that PBI should be considered an important tool to achieve universal health coverage of quality services. It was also noted that performance management and performance-based incentives should evolve to become the norm in the healthcare field globally, and implemented routinely across the levels of care. Reframing PBI for quality as a shared norm may help to move the field from targeted, pilot programs to scale up at a national level, while allowing room for creativity and experimentation according to the local context.

### ***Stakeholders***

The discussion that the group had around stakeholders noted that if PBIs were aligned with country priorities and had strong political commitment, there would be every chance for success. By not having a PBI initiative perceived as “donor driven,” but rather driven and managed by government stakeholders, they reasoned that PBI schemes could serve as a useful tool for the national government to signal priorities to state and local governments while, at the same time, allowing regional health care providers to make locally-relevant decisions. The group also recommended that all the different stakeholders – healthcare providers, policy makers, donors, governments, etc.--be brought together. Beyond government actors, they felt that civil society should play an important role to maintain accountability through mechanisms such as citizen monitoring.

### ***Sustainability and financing***

The group noted that domestic financing may be critical to sustainability and that national level funding is useful for increasing accountability and transparency. It also can be helpful to integrate financing from donors so that there is a plan for sustainability at the outset. Increasing decentralization of health systems promotes autonomy (as well as transparency) and verification and measurement can be an expensive, but necessary, part of this process. Strengthening current data collection systems so this effort is sustained in the long term can be viewed as an investment that will contribute to decreasing costs over time.

### ***Ethics***

Participants discussed the myriad ethical issues that may affect PBI programs and recommended that those running programs be vigilant about not inadvertently introducing discrimination or causing any

unintended ill effects. They also noted that the ethics of PBI programs warrant further exploration and that there should still be space for creativity and innovation within this framework. Civil society may be able to play multiple roles related to the implementation and experience of PBI programs in maternal newborn health, as well as other programs, and assist with adapting programs in response to routine monitoring and evaluation.

### ***Learning and sharing***

The small group discussed a variety of needs related to learning and sharing in real time, especially as a way to make sure that successes are capitalized on, successful programs replicated, and common pitfalls avoided. If this learning could be shared between and across countries, there might be the opportunity to catalyze new work as well. Importantly, the group identified several key issues for which sharing useful lessons would benefit: how data can be collected as well as utilized in real time; models for training healthcare workers; and identifying system readiness and pre-conditions for successful implementation of PBI. The discussion about ways to communicate this learning noted virtual discussions; dissemination of sharing via conversations, meetings, and conferences; and documenting and sharing knowledge informally as well as through formal channels, such as peer-reviewed literature.

The five recommendations that emerged from this discussion included:

### ***Recommendations***

- 1) Frame PBI as the mechanism to implement quality MNH services across entire systems as part of UHC.
- 2) Work with key stakeholders from government to own and implement PBI, making sure that there is space for local adaptation if/as needed, as well as engage civil society for accountability.
- 3) Consider mechanisms to integrate donor and national funding for PBI so that there is national-level ownership of the process (including data collecting) and outcomes and a greater chance of sustainability.
- 4) Monitor carefully the implementation and experience of PBI programs (perhaps with the assistance of civil society) in order to ensure there are no unintended negative effects and, if any are discovered, that course corrections can be made efficiently.
- 5) Implement multiple ways to share learnings from PBI programs in real time especially on key issues such as provider training; recognition of common pitfalls; identification of system readiness; and pre-conditions for successful implementation of PBI.

## **C. METRICS**

The small group discussing metrics noted that PBI programs measure a variety of quality indicators in facilities, including facility readiness, health workforce knowledge and adherence to clinical standards, and patient experiences. Programs employ various methods to measure quality indicators, and these methods have unique advantages and disadvantages and range widely in terms of complexity and the corresponding time and cost requirements. Two inter-related challenges were discussed:

1. How can programs get beyond structural indicators – which assess whether the inputs needed for high quality care are in place – to ***assess the quality of care that was actually delivered to patients*** through process and outcome indicators of quality?
2. How can programs determine the ***right mix of quality assessment indicators and methods*** that mitigates the time and resource costs of assessment without sacrificing accuracy.

For example, programs have been able to more easily assess the structural indicators of quality including health workforce quantity (i.e. how many trained providers are at each facility) and health facility readiness (i.e. the presence of necessary physical inputs to provide a minimum quality of care) through checklists and facility audits. Methods to assess the quality of care actually provided tend to be more resource-intensive. Direct observation is one method, but brings implications for patient privacy as well as bias due to the Hawthorne effect. Chart reviews are another method, but this relies on accurate and detailed record keeping within patient charts. A number of innovative alternatives have been tested in PBI impact evaluations as well as in country programs, including text-based and video vignettes and simulations. ***Learning from these recent experiences in quality measurement innovation and translating the findings for policy makers and programmers is essential*** as PBI designs are developed and adapted in countries.

The many methods to assess quality come at varying costs and yield results of varying accuracy or precision. In addition to the tradeoff between cost and accuracy, the group discussed other common pitfalls or drawbacks of current assessment methods, including the following:

- Some methods employed by PBI programs lend themselves to corruption and gaming the system. For example, self-assessment of provider-patient encounters by facility staff may be too burdensome to complete on the spot, so staff complete the assessments after-the-fact. Delayed completion can lead to inaccuracies and even exaggerations or fabrications. Systems must be in place to counteract these issues, for example, multiple layers of verification can help prevent gaming of the system.
- PBI programs may cause negative unintended consequences as a result of incentivizing only certain indicators or services. Facilities may focus their effort on the services that are “payable” and ignore other services or steps of a clinical process that are not incentivized or measured.
- The success of any PBI program relies on data systems at the country, district, and facility levels. Many countries implementing PBI programs have less than robust data systems, which may impose potential limitations on efficiency in PBI-based quality measurement. The challenge of integrating PBI data into country health information systems must be addressed.
- Providers have reported feeling overburdened with the data collection required to document quantity and quality of incentivized services, especially when they are already overburdened by a high patient volume.
- Third-party verification, which typically comes after the initial quality assessment is completed by facilities, presents several programmatic challenges, including avoiding conflicts of interest if verifiers are too closely associated with the facility (e.g. DHMTs). A colleague from the World Bank noted that the Bank invests heavily in verification and counter-verification and builds country capacity to do verification rather than fly in external verifiers. They also do risk-based identification of certain areas that need to be watched, rather than intensive verification of all facilities.
- They agreed that more work on cost-effectiveness of various quality assessment measures (and PBI programs more broadly) is needed to inform scale up.

In addition to the challenge of incorporating process quality measures, ***programs struggle to achieve an adequate yet manageable number of quality indicators*** in their quality assessment processes (i.e. provider competency and application of clinical standards). While PBI programs are able to assess facility

readiness through program checklists, the small group noted that programs must reconsider the value of including a vast number of structural indicators of quality (i.e. presence of supplies, pharmaceuticals, and other physical inputs) in every checklist. The PBI/QoC community would like to identify proxy indicators that can be used in place of many additional indicators, without losing accuracy. The group discussed the need for a more manageable list of quality indicators, as well as those most directly linked to health outcomes.

Another challenge that must be considered in designing quality assessment processes for PBI programs is how to incentivize providers to improve service quality without penalizing them in situations where ***structural deficiencies exist that are outside of providers' control***? In the presence of structural deficiencies in facilities, assessing provider knowledge may not be an accurate measure of adherence to standards during a clinical encounter with a patient. Providers may have knowledge of clinical standards and protocols, but cannot apply that knowledge due to constraints at the facility level such as shortages of vital pharmaceuticals or supplies, or understaffing. In such cases, a financial incentive for the provider is generally not sufficient to overcome the facility deficiencies preventing the provision of high-quality care.

In addition, the group agreed that ***PBI programs do not assess patient experiences of care well enough***. While household surveys or exit interviews enable assessment of patient experiences, issues of response bias must be addressed, and the costs of these methods are often prohibitive. Further, PBI programs rarely incorporate patient experience measures as part of the calculation of the quality score. Even though household surveys capture rich information on patient experiences of care, PBI programs often use this data solely to verify utilization of services as captured in patient registers. Patient experiences are often only considered in impact evaluations rather than in routine program monitoring. The group noted that programs should incorporate more and better measures of patient experiences of care into the quality checklists. They also called for household surveys to include questions for women on respectful and dignified care and noted that civil society organizations have a role to play in capturing the community and patient perspectives of facility-based health care.

The group noted that ***adapting PBI quality measures over time is critical***, including adjusting the weighting of different indicators, creating a tiered system, or removing unnecessary indicators. Routine review and revision has proven difficult to achieve in practice according to several country experiences, and support to country PBI programs should focus on this important priority. As part of this process, indicators can be honed to ensure they are assessing quality as intended. For example, one country found that incentivizing the weighing of the baby was not sufficient – it was more effective to incentivize the assessment of whether the baby's weight is sufficient.

Following the rich discussion about metrics and quality assessment, the group made the following five priority recommendations:

### ***Recommendations***

- 5) Establish a more robust learning community. Engage with colleagues regularly about successes and failures; nuances; learnings; and ways to catalyze progress in quality measurement.

- 6) Make manuals, checklists, and data publically available for the sake of transparency and learning; incorporate a participatory checklist revision process regularly. Work to evolve programs from mostly structural to process or outcome measures.
- 7) Identify ways to support regular review and revision of quality assessment within country programs.
- 8) Ensure issues of facility infrastructure, supply chain, and procurement processes, which affect quality but may be outside of provider/facility control, are taken into account in developing quality incentives.
- 9) Consider potential innovations in data generation and use, such as integrating data collection into existing information systems; investigating the issue of “weighting” data to clarify whether it leads to a valid quality index; and keeping the amount of data collected manageable so that it actually can be used and shared.

## SESSION 8: GALVANIZING LEADERSHIP

Jim Sherry from URC’s TRAction project moderated a panel with Fannie Kachale from the Malawi Ministry of Health; Moazzam Ali from WHO; and Ingvar Olsen from Norwegian Agency for Development (Norad) to wrap up the technical consultation on the impact of PBIs on the quality of maternal newborn care.

He started by asking an important internal question of the group, **“What do you want from this group that would make your work easier?”** The answers were clear. Kachale noted that countries need to be consulted from the beginning during the planning stages for these kinds of programs saying “...it needs to align with country and governmental priorities and should not be donor driven...” She went on to explain that members of the government need to be involved in the very first phase to ensure buy-in and that every effort has to be made to ensure that PBI schemes become part of the strategic health road map that the government already has.

Moazzam Ali explained that to get government ownership of PBI programs and support for the research that the World Bank is doing, in one country they held a meeting with the government to refine the question and solicit input for the intervention in the planning phase. The government was reluctant at first, he noted explaining “...they asked a lot of us...and we had to explain that the WB could only accomplish a limited set of objectives, not everything.” In this particular program, the government in question initially doubted that a voucher program could work, but the data convinced them it was applicable in their context and then they were convinced to earmark money in their budget.

Ingvar Olsen noted that the current context in many locations backs up this point. Traditional bilateral support has been reduced over time. There is tension between humanitarian work and development aid; budgets have been reduced. He explained that Norad saw an urgency to find better health solutions in countries. He noted, “We don’t want to push this [PBI]... on any country as the solution. However, we are eager to test it to see when it works and where it doesn’t work.” He added, “I doubt that PBF is a solution everywhere, but it has an important role to play.” Echoing points made by the other panelists, he noted that government leadership within countries is clearly the path towards sustainability.

Sherry then asked whether political will was sufficient, given that often countries have political will early in the process, and then begin to think about cost-effectiveness much later. He asked, “How much



should our research focus on PBI—or should it focus more on the bigger principles that affect the whole healthcare system?” The individuals in the panel agreed that PBI is a means to an end. Explained one, “It gives us a way to get into the larger health care system.” Another noted that what is important about PBI is that it serves as a way to move the system itself—in this case, especially the part of the system that addresses quality issues in maternal newborn health. “PBI lends us the opportunity to promote certain things, such as outcome-oriented issues, quality improvement, transparency, accountability. PBI can act as a catalyst to move the health system to be better functioning,” explained one of the panelists.

In further discussion of financing, Ministries of Health were noted as important agents of change. Given that many health systems are becoming decentralized, there is power in financing different levels of the health systems. Financing from federal to state, then state to districts, based in part on performance, was noted as an important tool. Additionally, one participant noted, it has the potential to change the relationship between the Ministry of Finance (MOF) and the Ministry of Health (MOH) since MOF often look at MOH as something they put money into without understanding what they get out of it. “Most Ministries of Health can’t even spend 100% of the budgets available to them,” noted one participant. Moving toward results is a good way to showcase to Ministries of Finance what the health sector is able to accomplish. Additionally, if there can be a lens on efficiency and equity, then we will all learn more about what matters to the entire system.

Importantly, a comment from Emma Iriarte from the Salud Mesoamérica Initiative noted that countries form one type of leadership, but the community represented in the room --donors, researchers, policymakers—have another and important type of leadership to offer. She posed a question about how we can be more effective in approaching countries working on PBI and quality noting that we all have different agendas, interests, and motivations. Olsen noted that we may be in the process of learning much more about these communities of leadership as we turn toward the Global Financing Facility and try to learn from the process of funding specific efforts in specific countries. Finally, Sherry noted, “One day, paying for coverage may well be the same as paying for a certain level of quality—or at least we can hope!”

## APPENDIX A: PROGRAM AGENDA

# Performance-based Incentives in Maternal Newborn Health

## *The role of quality*

### MEETING OBJECTIVES

- 1) To share the state of the evidence on whether and how performance-based incentives (PBI) affect quality of care, especially as applied to the field of maternal newborn health, from key low- and middle-income countries.
- 2) To understand existing indicators and methods for assessing quality (structural, adherence to clinical standard of care, patient experience) and the strengths, challenges, and considerations of each
- 3) To consider what we know about the intended and unintended consequences of PBI programs related to:
  - a. Health workforce
  - b. Quality of patient experience
  - c. Health facility readiness
- 4) To identify implementation research and quality measurement priorities to advance the effectiveness of PBI programs in addressing quality of care.
- 5) To consider how the state of the evidence can and should inform practice and policy moving forward for PBI interventions.

### DAY 1, WEDNESDAY, APRIL 20

#### *Morning*

*Key questions on performance-based incentives and their role in improving quality of care*

- 1) *What are the roles of incentives, both financial and nonfinancial, in the provision of high quality care?*
- 2) *How can quality be defined and measured?*
- 3) *What are the ethical considerations of using incentives to achieve increased utilization of services in settings where service quality is poor?*
- 4) *What is the evidence on the effect of PBI on quality and what are the evidence gaps?*

9:00-9:05	Welcome	Ana Langer, Harvard T.H. Chan School of Public Health
9:05-9:20	Rationale for meeting; overview of agenda and objectives; participants	Neal Brandes, USAID
9:20-10:00	Performance-based Incentive Programs: How and why do they work? When don't they work? (Moderated discussion)	Peter Berman, Harvard T.H. Chan School of Public Health Rena Eichler, Broad Branch Associates Sebastian Bauhoff, Center for Global Development
10:00-10:30	Who defines quality and why does it vary: A multi-country analysis <i>(Followed by discussion)</i>	Margaret Kruk, Harvard T.H. Chan School of Public Health Debbie Armbruster, USAID to lead discussion
10:30-11:00	The ethics of performance-based-incentives and quality frameworks: key questions and considerations <i>(Followed by discussion)</i>	Nir Eyal, Harvard T.H. Chan School of Public Health
11:00-11:30	<i>Coffee break</i>	
11:30-12:45	An overview of World Bank performance-based incentive programs addressing quality of care (coverage, quality and access) and what is known about evidence and impact <i>(Followed by discussion)</i>	Dinesh Nair, World Bank Damien de Walque, World Bank
12:45-1:45	<i>Lunch break</i>	

DAY 1, WEDNESDAY, APRIL 20		
<p><i>Afternoon</i>  <i>Key questions on performance-based incentives in the context of maternal newborn health</i></p> <ol style="list-style-type: none"> <li>1) <i>Where are PBI programs that address quality of care operating and what are the defining characteristics of these programs?</i></li> <li>2) <i>What can we learn from the range of quality assessment methods used in PBI programs?</i></li> <li>3) <i>What can we learn from country experiences?</i></li> </ol>		
1:45-2:15	Discussion on what we know and what our questions are based on morning sessions	Ingvar Olsen, NORAD
2:15-2:45	The landscape of PBI programs addressing quality of care and what we can learn from the quality assessment indicators employed	Jessica Gergen, Thinkwell Global
2:45-3:15	<i>Coffee break</i>	
3:15-5:00	Moderated panel discussion: Malawi's and Zimbabwe's experiences in PBI & quality of care <i>(Followed by large group discussion)</i>	Fannie Kachale, MOH/Malawi Stephan Brenner, University of Heidelberg/Malawi Son-Nam Nguyen, World Bank/Zimbabwe interviewed by Supriya Madhavan, USAID
5:00-5:15	Wrap up	Samantha Ski, URC/TRAction
5:15-6:30	Cocktail Reception	All are welcome

<b>DAY 2, THURSDAY, APRIL 21</b>		
<p><i>Morning</i></p> <p><i>Key questions</i></p> <ol style="list-style-type: none"> <li>1) <i>What can we learn from country experiences?</i></li> <li>2) <i>What questions does an analysis of the evidence raise?</i></li> <li>3) <i>What specific recommendations are relevant to move practice and policy forward?</i></li> <li>4) <i>How can leadership be galvanized on critical priorities?</i></li> </ol>		
9-9:15	Recap of Day 1's progress	Emma Iriarte, Salud Mesoamérica Initiative
9:15-10:15	Moderated panel discussion: Tanzania's, Senegal's and Mesoamerica's experiences in PBI and quality of care	Vincent Somville, Christian Michelsen Institute/Tanzania Sophie Faye, Abt Associates/Senegal Emma Iriarte, Salud Mesoamérica Initiative interviewed by Rifat Atun, Harvard T.H. Chan School of Public Health
10:15-10:45	Large group discussion on learnings shared from country experiences	Rifat Atun, Harvard T.H. Chan School of Public Health
10:45-11:15	<i>Coffee break</i>	
11:15-12:15	Discussion and analysis	<p>Small group 1: Evidence gaps and implementation science priorities led by Emily Peca, URC/TRAction</p> <p>Small group 2: Policy/governance and programs led by Bev Johnston, USAID</p> <p>Small group 3: Metrics led by Edward Broughton, URC/ASSIST</p>
12:15-12:45	Preparation of recommendations	Small group work
12:45-1:45	<i>Lunch break</i>	

DAY 2, THURSDAY, APRIL 21		
<p><i>Afternoon</i>                      Key recommendations on performance-based incentives in the context of maternal newborn health</p>		
1:45-2:15	Evidence gaps and implementation science recommendations + discussion	Presentation from group 1
2:15-2:45	Policy/governance and program recommendations + discussion	Presentation from group 2
2:45-3:15	Metrics recommendations + discussion	Presentation from group 3
3:15-3:45	<i>Coffee break</i>	
3:45-4:30	How can leadership be galvanized for PBIs and quality?  <i>(Moderated panel)</i>	Moazzam Ali, WHO Ingvar Olsen, NORAD Emma Iriarte, Salud Mesoamérica Initiative Fannie Kachale, Malawi Ministry of Health Moderated by Jim Sherry, URC/TRAction and CUNY
4:30-5:00	Wrap up	Ana Langer and Jim Sherry

## APPENDIX B: PARTICIPANT LIST

# Performance-based Incentives in Maternal Newborn Health

## *The role of quality*

**ADAMOU, Bridgit**

University of North Carolina at Chapel Hill  
MEASURE Evaluation Project  
[adamou@email.unc.edu](mailto:adamou@email.unc.edu)

**ALI, Moazzam**

World Health Organization  
[alimoa@who.int](mailto:alimoa@who.int)

**ARMBRUSTER, Debbie**

United States Agency for International Development  
[darmbruster@usaid.gov](mailto:darmbruster@usaid.gov)

**ATUN, Rifat**

Harvard T.H. Chan School of Public Health  
[ratun@hsph.harvard.edu](mailto:ratun@hsph.harvard.edu)

**BAUHOFF, Sebastian**

Center for Global Development  
[sbauhoff@cgdev.org](mailto:sbauhoff@cgdev.org)

**BERMAN, Peter**

Harvard T.H. Chan School of Public Health  
[pberman@hsph.harvard.edu](mailto:pberman@hsph.harvard.edu)

**BRANDES, Neal**

United States Agency for International Development  
[nbrandes@usaid.gov](mailto:nbrandes@usaid.gov)

**BRENNER, Stephan**

University of Heidelberg  
[stephan.brenner@uni-heidelberg.de](mailto:stephan.brenner@uni-heidelberg.de)

**BROUGHTON, Edward**

USAID|ASSIST Project

[ebroughton@urc-chs.com](mailto:ebroughton@urc-chs.com)

**COLE, Bethany**

Fistula Care Plus Project, EngenderHealth  
[bcole@engenderhealth.org](mailto:bcole@engenderhealth.org)

**de WALQUE, Damien**

World Bank  
[ddewalque@worldbank.org](mailto:ddewalque@worldbank.org)

**EICHLER, Rena**

Broad Branch Associates  
[renaeichler@broadbranch.org](mailto:renaeichler@broadbranch.org)

**EYAL, Nir**

Harvard T.H. Chan School of Public Health  
[neyal@hsph.harvard.edu](mailto:neyal@hsph.harvard.edu)

**FAYE, Sophie**

Abt Associates  
[sophie\\_faye@abtassoc.com](mailto:sophie_faye@abtassoc.com)

**GERGEN, Jessica**

Thinkwell Global  
[jgergen@thinkwell.global](mailto:jgergen@thinkwell.global)

**IRIARTE, Emma**

Salud Mesoamérica Initiative  
[emmai@iadb.org](mailto:emmai@iadb.org)

**JOHNSTON, Bev**

United States Agency for International Development  
[bjohnston@usaid.gov](mailto:bjohnston@usaid.gov)

**KACHALE, Fannie**

Ministry of Health, Malawi  
[fankachale@yahoo.co.uk](mailto:fankachale@yahoo.co.uk)

**KRUK, Margaret**

Harvard T.H. Chan School of Public Health  
[mkruk@hsph.harvard.edu](mailto:mkruk@hsph.harvard.edu)

**LANGER, Ana**

Harvard T.H. Chan School of Public Health  
[alanger@hsph.harvard.edu](mailto:alanger@hsph.harvard.edu)



**MADHAVAN, Supriya**

United States Agency for International Development  
[smadhavan@usaid.gov](mailto:smadhavan@usaid.gov)

**NAIR, Dinesh**

World Bank  
[dnair@worldbank.org](mailto:dnair@worldbank.org)

**NELSON, Jennifer**

Salud Mesoamérica Initiative  
[jennifern@iadb.org](mailto:jennifern@iadb.org)

**NGUYEN, Son-Nam**

World Bank  
[snguyen@worldbank.org](mailto:snguyen@worldbank.org)

**OFMAN, Julia**

Maternal Health Task Force  
[jofman@hsph.harvard.edu](mailto:jofman@hsph.harvard.edu)

**OLSEN, Ingvar Theo**

Norwegian Agency for Development Cooperation  
[ingvar.olsen@norad.no](mailto:ingvar.olsen@norad.no)

**PECA, Emily**

USAID|Translating Research into Action  
(TRAction) Project  
[epeca@urc-chs.com](mailto:epeca@urc-chs.com)

**POSTOLOVSKA, Ira**

Harvard T.H. Chan School of Public Health  
[iryana.postolovska@gmail.com](mailto:iryana.postolovska@gmail.com)

**SHERRY, Jim**

USAID|Translating Research into Action (TRAction) Project  
[jsherry@urc-chs.com](mailto:jsherry@urc-chs.com)

**SKI, Samantha**

USAID|Translating Research into Action (TRAction) Project  
[sski@urc-chs.com](mailto:sski@urc-chs.com)

**SOMVILLE, Vincent**

Chr. Michelsen Institute  
[vincent.somville@cmi.no](mailto:vincent.somville@cmi.no)

**SURETTE, Andrea**

USAID|Translating Research into Action (TRAction) Project  
[asurette@urc-chs.com](mailto:asurette@urc-chs.com)

**WEGNER, Mary Nell**  
Maternal Health Task Force  
[mnwegner@hsph.harvard.edu](mailto:mnwegner@hsph.harvard.edu)

**ZENG, Wu**  
Brandeis University  
[wuzengcn@brandeis.edu](mailto:wuzengcn@brandeis.edu)