



## STUDY PROTOCOL

# Characterizing performance improvement in primary care systems in Mesoamerica: A realist evaluation protocol

[version 1; peer review: 2 approved, 1 approved with reservations]

Wolfgang Munar <sup>1</sup>, Syed S. Wahid<sup>1</sup>, Leslie Curry<sup>2</sup>

<sup>1</sup>Milken Institute School of Public Health, George Washington University, Washington, DC, 20052, USA

<sup>2</sup>Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, 06520-8034, USA

**V1** First published: 03 Jan 2018, 2:1  
<https://doi.org/10.12688/gatesopenres.12782.1>  
 Latest published: 04 Oct 2018, 2:1  
<https://doi.org/10.12688/gatesopenres.12782.2>




## Abstract

**Background.** Improving performance of primary care systems in low- and middle-income countries (LMICs) may be a necessary condition for achievement of universal health coverage in the age of Sustainable Development Goals. The Salud Mesoamerica Initiative (SMI), a large-scale, multi-country program that uses supply-side financial incentives directed at the central-level of governments, and continuous, external evaluation of public, health sector performance to induce improvements in primary care performance in eight LMICs. This study protocol seeks to explain whether and how these interventions generate program effects in El Salvador and Honduras.

**Methods.** This study presents the protocol for a study that uses a realist evaluation approach to develop a preliminary program theory that hypothesizes the interactions between context, interventions and the mechanisms that trigger outcomes. The program theory was completed through a scoping review of relevant empirical, peer-reviewed and grey literature; a sense-making workshop with program stakeholders; and content analysis of key SMI documents. The study will use a multiple case-study design with embedded units with contrasting cases. We define as a case the two primary care systems of Honduras and El Salvador, each with different context characteristics. Data will be collected through in-depth interviews with program actors and stakeholders, documentary review, and non-participatory observation. Data analysis will use inductive and deductive approaches to identify causal patterns organized as 'context, mechanism, outcome' configurations. The findings will be triangulated with existing secondary, qualitative and quantitative data sources, and contrasted against relevant theoretical literature. The study will end with a refined program theory. Findings will be published following the guidelines generated by the Realist and Meta-


## Open Peer Review

Approval Status   

	1	2	3
<b>version 2</b> (revision) 04 Oct 2018			
<b>version 1</b> 03 Jan 2018	 view	 view	 view

1. **Daniel H. Kress**, RTI International, Seattle, USA

2. **Jean-Paul Dossou**, Centre de Recherche en Reproduction Humaine et en Démographie, CNHU/HKM, Cotonou, Benin  
 Institute of Tropical Medicine of Antwerp, Antwerp, Belgium

3. **Lisa R. Hirschhorn** , Northwestern University, Chicago, USA

Any reports and responses or comments on the article can be found at the end of the article.

narrative Evidence Syntheses study (RAMESES II). This study will be performed contemporaneously with SMI's mid-term stage of implementation. Of the methods described, the preliminary program theory has been completed. Data collection, analysis and synthesis remain to be completed.

## Keywords

El Salvador, Honduras, Primary Care accountability reforms, Primary Care Performance in low- and middle-income countries, Primary Care performance measurement, Realist Evaluation, Results-based financing, Salud Mesoamerica Initiative

**Corresponding author:** Wolfgang Munar ([wolfgangmunar@email.gwu.edu](mailto:wolfgangmunar@email.gwu.edu))

**Author roles:** **Munar W:** Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Wahid SS:** Conceptualization, Investigation, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Curry L:** Conceptualization, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work was supported by the Bill and Melinda Gates Foundation (grant number OPP1154415).

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2018 Munar W *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Munar W, Wahid SS and Curry L. **Characterizing performance improvement in primary care systems in Mesoamerica: A realist evaluation protocol [version 1; peer review: 2 approved, 1 approved with reservations]** Gates Open Research 2018, 2:1 <https://doi.org/10.12688/gatesopenres.12782.1>

**First published:** 03 Jan 2018, 2:1 <https://doi.org/10.12688/gatesopenres.12782.1>

## Introduction

Improving performance of primary care systems in low- and middle-income countries (LMICs) has been suggested as a necessary condition for the achievement of universal health coverage in the age of the Sustainable Development Goals<sup>1</sup>. High-performing primary care systems not only are the first point of contact for continuous, coordinated, comprehensive and people-centered health services, but also provide critical preparedness and response to public health threats<sup>2,3</sup>.

The Salud Mesoamerica Initiative (SMI) is a multi-country, large-scale primary care performance improvement program. It is the result of a partnership between the governments of the eight Mesoamerican nation-states, the Bill and Melinda Gates Foundation, the Carlos Slim Foundation, the Government of Canada, the Inter-American Development Bank (IADB) and, during earlier stages, the Government of Spain. The program is aimed at improving reproductive, maternal, neonatal and child health outcomes among the poorest, rural populations in participating countries. Intended outcomes include increased availability, utilization, and effective coverage of primary care services and a reduction in preventable health inequities. The program's approach to performance improvement combines the use of high-powered financial incentives at the government-level and the external verification of public sector, primary care system performance.

Programs and policies aimed at improving the performance of health systems have been at the forefront of many public-sector reforms. Initial waves of reforms in the public sectors of high-income countries were focused on learning and improvement<sup>4-7</sup>; subsequent waves of reform targeted public sector accountability and organizational best-practice<sup>8,9</sup>. Governments in LMICs adopted and replicated these reforms, oftentimes with the support of multilateral finance institutions and agencies in the official development assistance space.

Several generic types of reforms follow the logic of public-sector interventions<sup>10</sup>, including (1) political interventions as expressed in policies and regulations; (2) reforms by laws; (3) intervention by audit and inspection, based on continuous evaluation of results and conformity to predefined norms; (4) intervention by management, based on organizational science and management practice, such as continuous quality improvement or change management methods, among others; and, (5) intervention by rationalizing professional behaviors such as the introduction of evidence-based practices and the standard comparison of outcomes by public sector providers.

In LMIC, accountability-driven reforms have flourished under the rubric of Results-based Financing (RBF); the health sector has regularly been at the center of such reforms. The Multilateral Finance Institutions, the Global Fund to Fight HIV/AIDS, tuberculosis and malaria, and the Global Alliance for Immunizations

have financed results-oriented, global health programs, some of which have targeted health system improvements.

There are ambiguities in the definitions as well as in the scope and content of RBF programs and policies. In this study protocol, RBF is understood as “any program that rewards the delivery of one or more outputs or outcomes by one or more incentives, financial or otherwise, upon verification that the agreed-upon result has actually been delivered”<sup>11</sup>; incentives can target health care providers (supply side), households (demand side), or both. Performance-based Financing (PBF), is a prevalent type of RBF in which the incentives are exclusively financial; rewards are only aimed at providers; and the payments are usually adjusted for quality. PBF assumes many forms but, in essence, serves the purpose of reforming the ways in which governments pay health care providers (individuals and facilities) for the provision of services.

Accountability-driven interventions in public sector reforms are designed to reduce the misalignment in incentives between principals (voters, legislative bodies, executive-level leadership, funders, etc.) and their agents (program implementers, providers of care, etc.)<sup>12,13</sup>. Such reforms usually assume that incentives and rewards serve as powerful motivators for the achievement of desirable behaviors among utility-maximizing, rational individuals<sup>14-16</sup>. These assumptions have conventionally been based on principal-agent theory<sup>14,15</sup>, positive agency theory<sup>16</sup>, and/or rational choice theory<sup>17,18</sup>. In recent years, there have been calls for using a more expansive view of human agency when discussing motivation and decision-making. Under such views, humans are not exclusively motivated by rewards and incentives, but can also be driven to action by intrinsic motivators<sup>19</sup>. This perspective has also influenced contemporary research on PBF in LMIC<sup>20,21</sup>.

Most of the primary studies that have assessed RBF and PBF programs in LMIC have characterized the effects of financial incentives on provider-level motivation and behaviors<sup>21-31</sup>. However, the evidence on the effects of RBF on large-system reforms targeting government-level improvements is scarce; and studies in LMIC settings are largely absent<sup>32-34</sup>.

The most-recent systematic review of performance-based financing programs in LMIC concluded that PBF is not a single type of intervention and that its effects are dependent on the interactions among multiple variables<sup>31</sup>. Most PBF evaluations to date have used a narrow focus such as characterizing changes in health care outputs, while neglecting most other domains of primary care performance improvement<sup>35</sup>. Furthermore, the empirical evidence about how PBF leads to changes in attitudes and behaviors among public sector actors is scarce<sup>35</sup>. Domains that have been understudied include, among others, whether and how do extrinsic and/or intrinsic motivators affect the behaviors, autonomy and

responsiveness of providers and managers of primary care delivery systems; the influence that performance measurement data can have on the behaviors of primary care system actors and stakeholders; and the negative effects of RBF reforms, such as gaming, shirking and cream-skimming.

Public-sector reforms tend to incorporate multiple interventions that generate effects at different levels within organizational hierarchies, and among different actors and stakeholders. Those actors and the environments in which they are embedded interact with each other through time, generating inter-dependencies and, oftentimes, leading to counter-intuitive, emergent, and unintended effects. Furthermore, the implementation strategies and ancillary components of reform programs themselves, such as the provision of technical assistance or change management support, can also trigger system changes that need to be better studied<sup>31</sup>.

Beyond accountability reforms, studies on performance management and performance assessment have empirically studied improvement-driven public-sector reforms. Studies of such reforms in the public sector of the United States have identified factors that can drive organizational learning and improvement. For instance, Moynihan and Landuyt<sup>36</sup> found that the most influential predictors of organizational improvement and learning were the use of work-groups as learning forums; the availability of performance information systems that collect, store and disseminate performance data; the existence of a mission orientation that builds a sense of shared vision for success and common purpose; and the existence of organizational slack, such as time and resources that allow people to think and learn.

Improvement reforms are predicated on the assumption that the continuous collection, availability and analysis of performance data and information would lead to organizational improvement and learning<sup>7</sup>. However, despite widespread calls for using performance data and information to improve decision-making, the utilization of such data and information can rarely be guaranteed<sup>37</sup>. Also, little is known about the conditions under which performance measurement work or the mechanisms that lead to system improvement. Studies in evaluation science have addressed these issues<sup>38,39</sup>. In this literature, the availability and dissemination of performance evaluation can influence system improvement through multi-level changes on individual, inter-personal and/or collective motivation. Considerable research in evaluation science has been informed by this evidence<sup>40–44</sup>. We did not find, however, any study assessing the effects of evaluation results on health system performance improvement in LMICs.

## Study setting

In SMI, governments agree with the IADB to the implementation of up to three consecutive, 18–24 month programs, aimed at achieving a series of progressively complex health targets (including inputs, processes, outputs and outcomes) that are externally verified by the University of Washington's Institute for Health Metrics and Evaluation (IHME). Participating governments contribute domestic funds *a-priori* to attain the agreed-upon targets; once domestic funds are made available and targets

are agreed, SMI matches the domestic contribution with grant financing on a 1:1 ratio. Afterwards, the IADB enters into formal performance contracts with each government. In the contract, SMI commits to reimbursing half of the initially invested domestic funds contingent on the achievement of the agreed-upon performance targets<sup>45</sup>.

Country-specific performance frameworks with geographical targeting of the poorest, rural populations, were negotiated with each government at the start of the program and have remained stable through time. A pass-or-fail policy was agreed, according to which a government has to achieve 80% or more of the approximately ten (10) targets that make-up any given performance framework to be eligible for the reimbursement of half of the initial domestic contribution. Table 1 lists some of the targets agreed by El Salvador and Honduras.

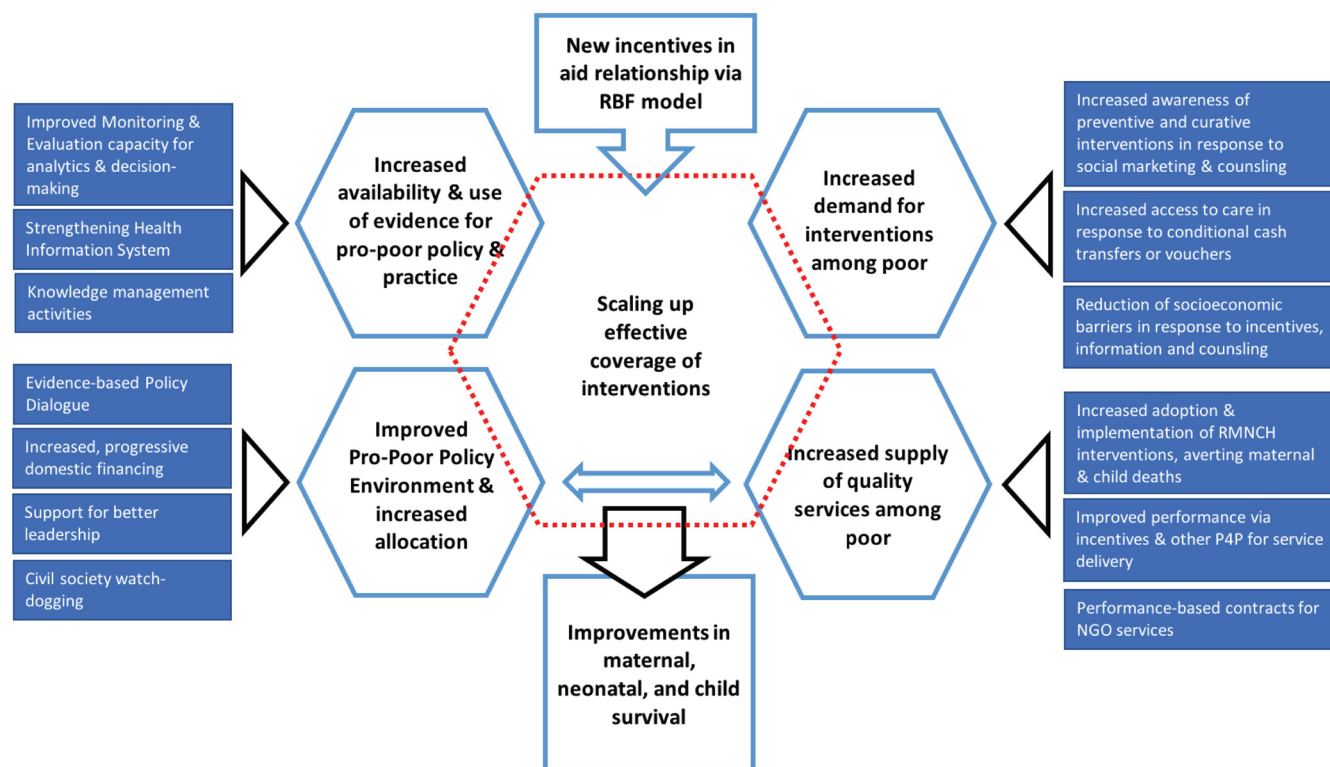
As the agency in charge of external verification of government performance, IHME conducts a full-scale quantitative measurement that follows SMI's sequential process of implementation. Before each country program starts, a baseline is collected and its results disseminated. After that, at the end of each 18 to 24-month implementation projects, IHME collects household and facility-based data to evaluate the achievement of agreed-upon results. Phase 1 programs started in a staggered fashion in 2011; phase-2 programs will finish during 2017; a third and final phase will start in 2018 and go into 2020. Program targets during phase 1 were focused on adherence to protocols, availability of resources and, in general, structure and process performance. During phase 2, targets were focused on outputs, and phase 3 will be centered on health outcomes, including but not limited to, coverage of exclusive breastfeeding, increased modern contraceptive prevalence rates, effective coverage of antenatal care and institutional deliveries, post-partum and post-natal care coverage, and in some cases, reductions in the prevalence of anemia and effective coverage of measles vaccination, measured in blood<sup>46</sup>. After each round of performance evaluation, results are aggregated and disseminated in each country through policy dialogue workshops convened by the government and involving the IADB and IHME.

SMI's original theory of change (Figure 1) hypothesized that the use of supply-side financial incentives directed to central-level ministries in each participating government (Ministries of Finance and Health) would focus their attention on accounting for the achievement of their own agreed-upon health targets. The success of this hypothesis rested on four causal pathways. The first established that the three consecutive, biannual rounds of external verification of performance by IHME would generate sustained pressure on governments for the production of health results. The second pathway proposed that ongoing dialogic, participatory dissemination of data, information and evidence would lead to progressive improvements in the quality of care services and improved, aggregate performance in each participating country's primary care system. Anticipated population-level health effects were also contingent on increasing domestic pro-poor health spending and expanding the demand for high impact health interventions among beneficiary populations.

**Table 1. Summary of performance frameworks in El Salvador and Honduras.**

Indicators	EL SALVADOR	Baseline	Target	Indicators	HONDURAS	Baseline	Target
	<b>First Phase</b>				<b>First Phase</b>		
Number of families enrolled in Family Health Teams		14,681	38,661	Health centers with permanent availability of micronutrient powder for supplementation at home		0	80%
Number of community health units with supply of four modern family planning methods (injectable, barrier, oral and intra-uterine devices).		11	65	Primary and second care level health units supplied with family planning methods according to ministry of health's current standard		86.4	90%
Review of national policy for micronutrient products distribution to children aged 6–23 months		No	Yes	Maternal & Child health clinics with permanent availability of medications and inputs necessary for treatment of obstetric and neonatal emergency		62.5	80%
Inclusion in the standard on proper therapeutic dosage of zinc for diarrhea treatment in children under five (20 mg of zinc for 10–14 days with each episode).		No	Yes	Second level health care units with permanent availability of medications, inputs and equipment necessary for treatment of obstetric and neonatal emergency		0	2
Percentage of pregnant women enrolled in the prenatal register who had a prenatal checkup with a physician or nurse before week 12 of pregnancy.		67	77	Maternal deaths reported and investigated according to standards in 2013		N. A.	80%
	<b>Second Phase</b>				<b>Second Phase</b>		
Percentage of women of childbearing age (15–49) currently using (or whose partner uses) a modern contraceptive method.		53.5	60.5	Women (aged 15–49) who received at least four prenatal checkups according to best practices by qualified personnel during their most recent pregnancy in the last 2 years		23.7	33.7
Percentage of women of childbearing age (15–49) who had a prenatal checkup according to best practices with a physician or nurse before week 12 in their most recent pregnancy		47.5	62.5	Women (aged 15–49) whose most recent delivery was attended by qualified personnel in a health unit in the last 2 years		68.6	76.6
Percentage of children aged 6–23 months who had a hemoglobin value of < 110 g/L. (Prevalence of anemia in children aged 6–23 months)		46.5	36.5	Neonates with complications (prematurity, low birth weight, asphyxia and sepsis) managed according to hospital standards in the previous two years		6.9	36.9
Percentage of mothers who gave their children (aged 0–59 months) oral rehydration salts and zinc in the last episode of diarrhea		4.4	24.4	Women with obstetric complication (sepsis, hemorrhage and eclampsia) managed according to national standards in their most recent delivery in the last two years		11	51
Percentage of women of childbearing age (15–49 years) whose most recent delivery was attended by trained personnel in a health unit in the last two years.		86.2	94.2	Mothers who report giving their children aged 6–23 months at least 50 packets of micronutrient powder in the last six months (36m)		0.1	15.1
	<b>Third Phase</b>				<b>Third Phase</b>		
Pregnant women treated at health centers in the last year who had at least one preconception consultation with quality in the year before their pregnancy.		-1	10	Women (aged 15–49 years) who currently use (or whose partner uses) a modern family planning method		66.8	76.8
Percentage of women of childbearing age (15–49 years) currently using (or whose partner uses) a modern contraceptive method.		-1	7 PP	Women (aged 15–49 years) whose most recent delivery was attended by qualified personnel in a health unit in the last two years		68.6	8PP
Women who received postpartum contraceptives in the last year.		-1	15PP	Newborns who received neonatal care within 3 days following birth according to standard in the last two years		-2	8PP
Women with obstetric complication (pre-eclampsia with severe symptoms, hemorrhage and sepsis) treated according to national standard.		-1	25 PP	Women with obstetric complication (sepsis, hemorrhage and eclampsia) managed according to the standard in their most recent pregnancy in the last two years		-2	25PP
Neonates with complications (low birthweight, prematurity, asphyxia and sepsis) treated according to the standard.		-1	25 PP	Neonates with complications (prematurity, low birth weight, asphyxia and sepsis) managed according to hospital level standards in the previous two years		-2	25 PP
Newborns who received neonatal care after birth according to the standard in the last two years.		-1	80%	Prevalence of anemia in children aged 6–23 months (Children aged 6–23 months with hemoglobin levels < 110 g/L)		35.3	25.3





**Figure 1.** SMI initial theory of change.

While the program's original theory of change identified several causal pathways, it did not explain *how* its main interventions would trigger outcomes, nor did it provide, either, *a-priori* explanations about the role that each country's policy context could play in moderating the effects of program interventions. The SMI partnership appears to have embraced a high degree of flexibility in implementation to facilitate governmental buy-in. In 2011, the partners agreed on a set of common principles including a focus on external and independent measurement of results, accountability and transparency, and country ownership. These principles established the institutional boundaries that, in turn, allowed the IADB to negotiate country-specific performance contracts, results frameworks and evaluation plans with each participating government. They also granted implementing partners a high degree of flexibility in the design of each country's multi-phased implementation plans and, also, led to performance contracts based on a few high-order principles (country ownership, a focus on results, pro-equity, cost-effective interventions, measurability and transparency) that were originally agreed among the funders and the IADB, and reflected in [the program's operating model](#).

In the two countries under study, El Salvador and Honduras, the program's focus on country ownership led to each government deciding how to best deploy SMI's non-reimbursable resources and their own domestic financing to increase the likelihood of achieving programmatic success. For instance, El Salvador had undergone a health system reform in the early 2010s, which

coincided with the beginning of SMI implementation. The government decided to focus its targets on results that leveraged one of the reform's central tenets, the provision of universal primary care services through community-centered, Family Health Teams<sup>47</sup>. Honduras, in turn, had started in the late 2000s a large-scale pilot of contracting-out and pay-for-performance in the delivery of primary care services<sup>48</sup>. The government thus decided to leverage its own performance-driven policies and programs and has thus implemented SMI in municipalities that had already acquired experience with RBF.

## Methods

### Methodological approach

The evidence gaps identified in our literature review have led to recent calls for new approaches in the evaluation of complex public-sector reforms, such as PBF<sup>30,35,49,50</sup>. It has been argued that realist evaluation provides a valuable and relevant approach for assessing interventions that involve changing human decisions and actions<sup>51,52</sup>. Realist evaluation is a form of theory-driven inquiry based on the premise an evaluation needs to answer "*what worked, how, in what circumstances and for whom*", rather than the conventional question "*Did the program work?*"<sup>52</sup>. The appeal of this approach, compared to other theory-driven methods, lies in its explicit foundations in critical realism – an epistemology located between positivism and relativism<sup>52</sup>. Such perspective contends that program interventions bring about social change through underlying, usually hidden causal mechanisms, and considers the role of context as indispensable in explaining causality.

This study addresses two research questions: (1) What are the effects of using supply-side financial incentives on the performance of the primary care systems in Honduras and El Salvador? How are those effects produced? Under what contextual factors are these effects produced in each country? And, (2) What are the effects of continuous external verification of performance in the two countries under study? How are those effects produced? Under what contextual factors are these effects produced in each country?

While there is no single way to implement a realist evaluation study, as the experience with its use and applications grow, various authors have adopted and adapted Pawson and Tilley’s approach and identified a series of steps that are described below<sup>52–57</sup>.

**Developing a preliminary program theory.** Realist evaluation starts with the development of a program theory that serves as a hypothesis about the ways in which outcomes are produced through the interaction between interventions and context conditions, and mediated by hidden, not-observable mechanisms. The latter have been defined as the ideas and opportunities triggered among program actors and stakeholders in response to program interventions<sup>57</sup>. The process of testing and refining program theories usually relies upon quantitative and qualitative methods and culminates with a refined program theory<sup>53</sup>.

This stage in the study was completed through complementary approaches, including (1) review of program design documents; (2) discussions with program designers to gain in-depth understanding of the original causal links between program interventions and expected outcomes; (3) scoping review of the literature, focused on identifying theories and empirical evidence addressing similar processes of primary care system change; and,

(4) facilitation of a workshop with IADB stakeholders, which helped understand their assumptions about how program interventions effects could be produced in the two countries under study. This process of making explicit the assumptions held by program stakeholders before data is collected is an essential aspect in realist evaluation. As a result of these various activities, the research team formulated a preliminary program theory.

In a separate, ongoing study we will perform a realist synthesis of performance improvement; performance measurement/evaluation; and, results- and performance-based financing. The process started with a scoping review of social science theories related to the two research questions and led to the mapping and synthesis of theories explaining the contextual factors and causal mechanisms of relevance to this study protocol. This will be followed by a search for primary studies, systematic reviews, realist evaluations and realist syntheses on the themes above and as required by our research questions. The search for the scoping review was done on Science Direct, JSTOR, and Google Scholar using a snowballing technique. The theories that were mapped are summarized in [Supplementary File 1](#).

Based on a synthesis of the results from the scoping review, and informed by the knowledge acquired from the stakeholder workshop and the document review, a preliminary program theory was developed ([Figure 2](#)). It is summarized below as a series of inter-linked propositions:

- The use of (1) high-powered, supply-side financial incentives aimed at central-level government actors and stakeholders (intervention 1) and the implementation of continuous, external evaluation and verification of

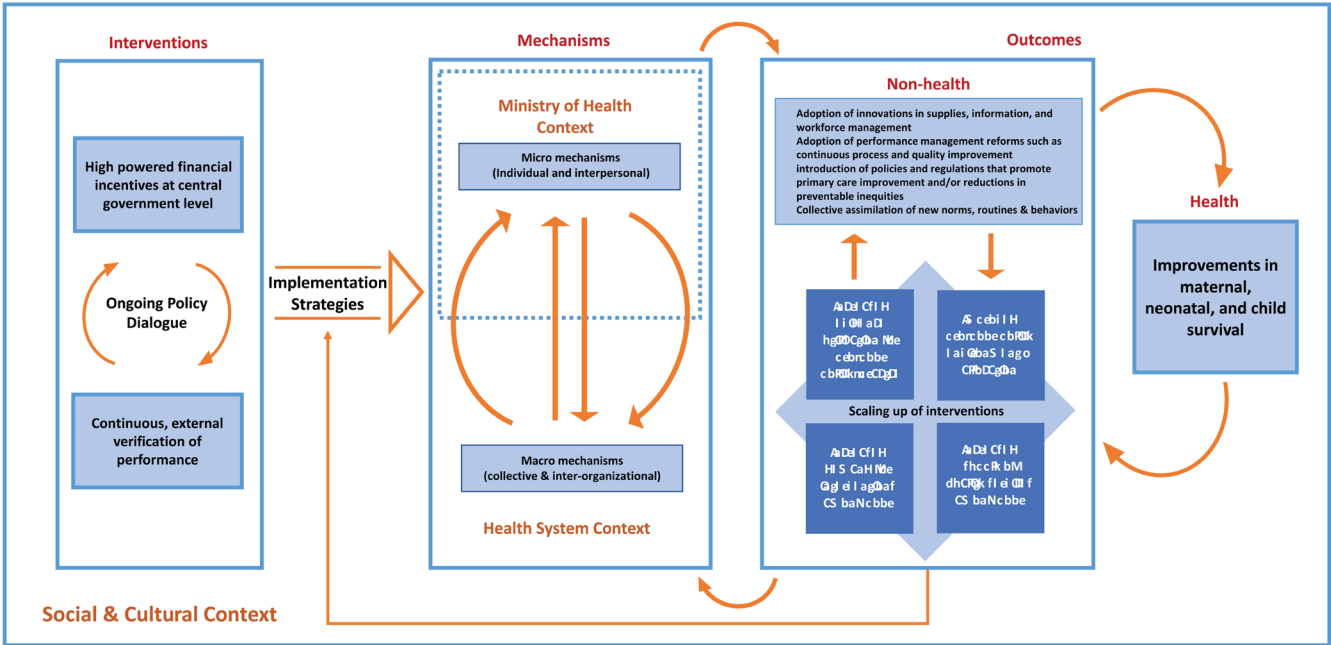


Figure 2. Preliminary program theory.

primary care service performance (intervention 2) supports country priorities through continuous policy dialogue, technical support, and purposive dissemination of performance results (implementation strategy);

- Leading to the adoption of innovations in supplies, information, and workforce management (outcome 1); the adoption of performance management reforms such as continuous process and quality improvement (outcome 2); the introduction of policies and regulations that promote primary care improvement and/or reductions in preventable inequities (outcome 3); and, improved, population-level health outputs and outcomes (outcome 4).
- The behavioral changes listed above occur at various levels within the primary care system, as follows;
  - o At the individual level, they satisfy psychological needs such as autonomy, competence and relatedness and/or the need to upgrade or improve personal goals and self-efficacy (individual-level mechanisms)<sup>58–63</sup>;
  - o At the interpersonal level, because of the aggregate internalization by multiple individual actors and stakeholders, of changes in ideas and opportunities; and/or through a growing sense of public service and/or community service (individual and interpersonal mechanisms)<sup>19,39</sup>.
  - o Collective level changes could also be triggered whereby the ideas and opportunities of a sufficiently large number of individual actors internalize or assimilate new norms, routines and behaviors which, in turn, spread across inter-organizational and social networks<sup>64,65</sup>, leading to the emergence of new organizational culture and collective behavior (outcome);
- Collective inter-organizational-level changes may further lead to the institutionalization and collective assimilation of aggregate individual- and interpersonal-level behaviors through imitation and/or the adoption of new professional and cultural norms, and/or innovative, pro-performance policies (outcome)<sup>33</sup> thus, increasing the likelihood of the production of population-level health effects (outcome) and, potentially, transforming the primary care system in a sustained fashion (outcome)<sup>32,33</sup>.

Global, institutional, and organizational contextual conditions are also needed for the attainment of program outcomes and for the triggering of the above mechanisms. They include, at the global and sub-regional levels, the existence of favorable conditions such as influential issue-specific global agendas that match existing governmental priorities or a history of interactions between national health agencies and their agendas, and between those and official development aid agencies and their agendas<sup>66–69</sup>. At the country-level, the availability of solid institutional environments (laws, regulations, ongoing public-sector reforms, etc.) can create windows of opportunity for the introduction of policy innovations

and, also, facilitate convergence between domestic policies and programs, and the externally-funded interventions. Finally, pre-existing environmental conditions, such as the organizational capacity to absorb new knowledge or the presence of climates that support and enable change, have also been associated with increased assimilation of service innovations<sup>70,71</sup> and need to be considered in the characterization of context.

**Study design.** In this step the preliminary program theory will be tested, further developed, and validated or rejected. A multiple case-study design with embedded units with contrasting cases was selected<sup>72</sup>. The contrasting case approach aligns well with the proposition that the two different country contexts in Honduras and El Salvador can trigger to-be-identified mechanisms that generate program results. Given the system-wide and reinforcing effects of the two interventions under study, we define each country's primary care system as the unit of analysis. Within each country, at least two high-performing municipal-level primary care delivery systems will be analyzed as embedded sub-cases, each with its unique contextual and service delivery structure.

This evaluation is an 18-month study running from May 2017 to December 2018, and executed contemporaneously with SMI's mid-term stage of implementation. The study seeks to maximize diversity in institutional and policy context to increase the likelihood of identifying variations in policy and program conditions and characterizing the process of change generated to date by the program in one low- and one middle-income country, respectively Honduras and El Salvador. Both countries have to date been exemplars of high-performance, which in SMI is defined as the continuous achievement of 80% or more of the targets agreed between each government and the IADB, and externally verified by IHME.

In each country, the study will assess the context, interventions, implementation approaches, and program effects, intended and otherwise. At the central level of government, the study will characterize program antecedents, policy and organizational context, primary care system's stewardship and policy-setting, and overall program management and implementation. At the local, municipal level, it will explore primary care delivery through Family Health Teams in El Salvador, and on public as well as non-profit, pay-for-performance providers. Primary data collection will include the methods described below.

**Data collection methods.** Realist evaluation is method neutral, and the nature of the research, the evaluation questions and the preliminary program theory determine the choice of study design and methodology<sup>52,57</sup>. The primary data collection methods to be used in this study include in-depth interviews, non-participatory observation, and document review.

In-depth interviews with key informants in each country will be conducted to identify individual, inter-personal and collective or organizational factors that may affect primary care system performance in each country under study. These interviews will also be used to elicit contextual elements that could act as barriers or facilitators for the delivery of SMI's interventions.



In this study, we aim to gain a high-level understanding of the causal mechanisms and pathways of performance improvement, as reflected in the preliminary program theory. SMI intervenes at the central as well as local levels of the primary care system, generating hypothesized feedback effects between both. The evaluation aims require the characterization of the interactions and inter-dependencies that occur among multiple actors in the primary care system; this would allow resulting data to help explain the complex nature of the process of performance improvement, and ultimately, help the team validate or revise the preliminary program theory. Accordingly, in-depth key informant interviews will be conducted with four sets of actors: (1) Country policy- and program implementation actors in Honduras and El Salvador; (2) Health care providers at primary care facilities in Honduras and El Salvador; (3) Performance verification and evaluation stakeholders at IHME; and, (4) Program designers at the IADB. Key informants will be recruited using a purposeful sampling approach<sup>73</sup>. Subsequently, the sample will be snowballed from the initial set of informants.

The study's sample size cannot be determined *a-priori*, but we expect to conduct approximately 80 key informant interviews, which will ultimately be determined based on theoretical saturation<sup>74</sup>. Respondents will be invited to participate voluntarily in the study; no compensation will be provided for participation. Interview guides will be used to conduct in-depth interviews; a series of probes will also be developed *a-priori* (Supplementary File 2). Interviews with country actors and stakeholders will be conducted in Spanish by bilingual members of the research team; IADB and IHME respondents will be interviewed in English. All interviews will be recorded and transcribed verbatim and, when applicable, professionally translated into English.

To document the process of policy dialogue, the study will use non-participant observation during the dissemination of the external verification of performance for the second phase of the program, in early 2018. The research team will document the process followed in the policy dialogue session, the agenda, components and intended objectives, the sequence of events that transpire following the results, and the reactions and actions by country actors and stakeholders. Summary memos of the observations will be generated to be maintained in the project files.

To further understand policy and program context, the study will review key program documents pertinent to the design, implementation and evaluation of SMI interventions in El Salvador and Honduras. Specific attention will be given to documenting the policy and program context in each country, identifying the implementation strategies in each country, assessing performance and evaluation frameworks, and identifying secondary data sources that could be used for further triangulation during the data analysis stage. A complete list of reviewed documents will be maintained, and included as a supplemental file with the final report of findings.

### Data analysis

Data analysis of the in-depth key informant interviews will be conducted using an integrative methodology that merges both inductive and deductive approaches<sup>74</sup>. We will construct a set of

*a-priori* codes drawing from the realist evaluation context, intervention, mechanism, and outcome structure, relevant theoretical literature domains, the stakeholder workshop, and the document review described above. This will be combined with emergent inductive codes identified from a rigorous open coding process.

In an initial stage of data analysis, two coders will analyze a subset of transcripts in an iterative and systematic manner using the constant comparison method, and afterwards finalize the codebook through negotiation<sup>75</sup>. Subsequent transcripts will be coded by three experienced coders using the final codebook.

The coded data will be appraised using two complementary analytic approaches. The research team will use iterative conceptual and pattern coding to identify major emergent inductive themes. At the conclusion of the process, the codes will be arranged into the four major categories of context, intervention, mechanism, and outcomes. The team will scan *within* each category, "vertically", to identify commonalities and thematic elements, e.g. multiple combinations of contexts that could facilitate/inhibit the interventions; or a confluence of interventions that are catalytic and reinforce one another, etc. Furthermore, the data will be analyzed *across* categories, or "horizontally," to identify causal patterns whereby certain outcomes are interrelated to program interventions that trigger mechanisms among primary care system actors under specific contextual conditions. We expect these two analytic approaches to be complementary, and to allow building context-mechanism- outcome (CMO) configurations that will then be gauged to determine which patterns plausibly explain how each intervention generated the observed effects, expected and otherwise. The final thematic structure will be used to refine the preliminary program theory. Data analysis will be done with nVivo Version 11 for Mac.

Evaluation results will be completed by integrating findings from the different data collection methods (interviews, notes from non-participatory observations, and secondary document analysis) to confirm, reject or further develop the preliminary program theory and the causal patterns identified. The findings will also be contrasted with secondary quantitative and qualitative data sets collected by IHME and others and with social science literature in search for mechanism-oriented theory that may provide explanation for the emerging causal patterns. The results will be a series of CMO configurations that are backed up by the empirical data that provide plausible causal explanations for the observed findings.

**Synthesis and refined program theory.** In this step the research team will link the emergent CMO configurations to the preliminary program theory, leading to the adoption, modification, or rejection of the preliminary program theory and will, then, formulate plausible explanations of how and why high-powered, supply side incentives and external verification of performance generate the observed results. The resulting explanations for the observed program effects will then be compiled in the form of narrative summaries, tables, and/or causal loop diagrams. The end of the study product is a final, refined program theory.

Study findings will be published in peer-reviewed periodicals and disseminated locally among policy-makers in the two countries to

be studied. The presentation of findings will be made following the Realist And Meta-narrative Evidence Syntheses study (RAMESES II) that was designed to provide guidance on quality assurance and uniform reporting and improve quality and consistency in the reporting of realist evaluations<sup>76,77</sup>.

### Quality control

A set of measures will be taken to increase the validity of the study in terms of reflexivity, credibility and confirmability, and enhance the trustworthiness, transparency, and accountability of the research<sup>78</sup>. All researchers will engage in the introspective practice of maintaining ‘personal biases memos’ to make explicit all self-identified biases and pre-conceptions that may effect the research process<sup>78</sup>. All analytic decision notes and memos, biases memos, document analysis syntheses, interview guides, research team meeting agendas and minutes, and analysis outputs including coded transcripts, conceptual frameworks, tables, etc. will be preserved to provide a verifiable audit trail.

### Discussion

The refined program theory and CMO configurations resulting from this study have several anticipated uses and applications. Program implementers like the IADB and Salvadorian and Honduran government actors, for instance, can use the findings to consider introducing adjustments in SMI’s implementation during its third and final phase (2018–2020). Also, given the study’s focus on exploring the linkages between ongoing, pre-existing policy mandates and priorities, it is plausible to expect study findings to be of relevance to further improve the evaluation and subsequent re-design of domestic health policies in El Salvador and Honduras. Program evaluators like IHME and other research groups, can use program findings to enhance ongoing evaluation activities or to inform the design of new evaluations that deepen one or more of the various casual patterns identified. For example, our research team intends to use the emerging CMO configurations in the area of performance management to inform the design of a new study to explore whether and how SMI quality improvement interventions produce gains in primary care performance. We also expect program findings to set the stage for further realist evaluations in other large-scale primary care reform in contexts other than Mesoamerica.

Another source of complexity in this study arises from the significant evidence gaps that we identified and from the multiple fields that would need to be rigorously studied to properly address the various outcomes generated by accountability and performance management reforms. As discussed before, such outcomes can occur at different levels of analysis (individual, organizational and collective) and in different contexts (high- as well as LMIC). Not only does this type of research demand inter- and multi-disciplinary capabilities within research teams, but it also calls for rigorous, systematic assessment and mapping of the evidence gaps<sup>79,80</sup>. Theory-based program evaluations of primary care performance improvement would also benefit from the publication of realist syntheses that rigorously appraise the literature in search of context-mechanism-outcomes and program theory<sup>81,82</sup>. Such studies would not only facilitate the work of

research teams currently addressing primary care performance improvement research, but would also strategically shape future health system research agendas, particularly in LMIC.

The research team has faced several challenges in shaping this study’s hypotheses, or preliminary program theory. Many of these challenges are common to other realist evaluations and have been discussed elsewhere<sup>53,83</sup>. One such challenge pertains to settling on an unambiguous and precise definition for what constitutes a mechanism. Several definitions in the literature are of a descriptive nature and focus on well-known features of mechanisms such as them being unobservable, context-specific and being able to generate effects. We settled on a definition of mechanisms as the ideas and opportunities triggered among program actors and stakeholders in response to program interventions<sup>57</sup>. Such an approach is consistent with a view of social change according to which the beliefs, choices and opportunities of individual actors and the interactions among them (micro-level) are the main drivers of social change. This approach also recognizes that the “macro,” social and cultural environment in which these individual actors are embedded can shape social change by means of the internalization of collective values, norms and institutions among individual actors<sup>84,85</sup>.

Based on these considerations, this study aims to, first, explore plausible causal explanations based on individual or group-based ideas and opportunities among program actors and stakeholders and, second, to ground those observations on an understanding of the policy and program context in which those actors and stakeholders are embedded. Therefore, we expect that any explanation of primary care system performance improvement needs to address both individual, micro-level, as well as collective, macro-level properties that “are not meaningfully attributed to individuals”<sup>84</sup>. Three specific types of mechanisms that explain social change are thus of interest to this study<sup>86</sup>.

*Situational mechanisms* refer to the macro, organizational-level environment in which SMI actors and their social interactions or linkages occur, including domestic policy-makers, ministry of health managers and primary care providers, among others. It also includes SMI stakeholders such as the implementation agency (the IADB) and the external evaluators of performance (IHME). This type of mechanism operates in the direction from macro environment to individual actors (macro-to-micro change). *Action-formation mechanisms* are those that explain how actors’ ideas and opportunities influence individual behaviors across the primary care system. In this type of mechanism, the interaction between program interventions and context, trigger changes in individuals’ ideas and opportunities that further influence others in the same social system. This type of mechanism can generate effects that spread from an individual actor to additional actors (micro to micro change). Finally, *transformational mechanisms* provide explanations of how the sum of new behaviors of multiple individual actors in the primary care system bring about change across the entities that conform the primary care system’s macro environment such as norms and institutions (micro-to-macro social change).

Due to limitations in scope, this study can only plausibly characterize some of the “macro” and “micro” mechanisms triggered during SMI’s initial stages up until its current, mid-term stage (2011–2017). It is also plausible to characterize intended and unintended effects generated during this same period. Findings can also be used identify propositions about downstream effects that could occur or not during the final implementation phase (2018–2020) and about the mechanisms that could help sustain desirable effects after implementation ends. Longer-term, transformational mechanisms, their anticipated effects and the underlying context-mechanism-outcome configurations will, however, remain outside our scope of work.

Another challenge in this study refers to the contested nature of the current definitions used to characterize the interventions that conform RBF, PBF or any of the various reforms that use supply-side incentives to drive accountability in public sector actors and, as is the case in SMI, across the entire primary care system. Like others before us, we settled on the definitions provided by Musgrove<sup>11</sup>, but we remain cognizant of the fact that PBF is not a single intervention and that its “ancillary” components can themselves generate system effects<sup>31</sup>. In this respect, this study frames SMI interventions as generic types of public-sector reforms aimed at inducing accountability and organizational improvement and learning. Given that the challenges in defining what these large-scale reforms contain in specific contexts –both in high and well as less-developed nations– the use of a realist evaluation approach will likely contribute to the theorizing of how and why specific contexts generate health and non-health effects in primary care performance management reforms.

Finally, given the method-neutrality that is central to realist evaluations, this study also faced the challenge of settling on a final sequence and content of research methods and activities. We decided to, first, follow the steps described by Vareilles and Marchal in relatively similar realist evaluations and studies<sup>53,57</sup>, but also relied on the Realist and Meta-narrative Evidence Syntheses study that provides guidance on how to improve quality and

consistency in the reporting of realist evaluations<sup>77</sup>. By aligning protocol design to these guidelines, we expect that the future publication of this study’s findings will adhere to current best practice.

### Ethical statement

The study’s protocol was reviewed and declared exempt by the George Washington University’s Institutional Review Board (study number 041733). The Ministries of Health of El Salvador and Honduras were informed of the proposed research by the IADB and provided written approval for the research activities.

Ethical approval documentation will be made available on request. The study will employ scrupulous adherence to the highest ethical standards, and current international and local legislation pertaining to research governance. The data collection will operate under explicit informed consent, which will be preserved in study records. Respondents will be given the choice to provide consent verbally on tape before the interviews, or in writing. To maintain anonymity, respondents will reserve the right to review the study outputs and withdraw consent if necessary. All identifying information will be removed from transcripts and stored separately with access restricted to the research team. All transcripts will be stored electronically in password protected cloud services, and physical documents will be securely stored at George Washington University, Milken Institute School of Public Health.

---

### Competing interests

No competing interests were declared.

### Grant information

This work was supported by the Bill and Melinda Gates Foundation (grant number OPP1154415).

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

### Supplementary material

Supplementary File 1: Social science theories mapped in the scoping literature review.

[Click here to access the data.](#)

Supplementary File 2: In-depth interview guides.

[Click here to access the data.](#)

## References

1. Bitton A, Ratcliffe HL, Veillard JH, *et al.*: **Primary Health Care as a Foundation for Strengthening Health Systems in Low- and Middle-Income Countries.** *J Gen Intern Med.* 2017; **32**(5): 566–71.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
2. Kruk ME, Porignon D, Rockers PC, *et al.*: **The contribution of primary care to health and health systems in low- and middle-income countries: a critical review of major primary care initiatives.** *Soc Sci Med.* 2010; **70**(6): 904–11.  
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Gates B: **The Next Epidemic—Lessons from Ebola.** *N Engl J Med.* 2015; **372**(15): 1381–4.  
[PubMed Abstract](#) | [Publisher Full Text](#)
4. Moynihan DP: **Managing for Results in State Government: Evaluating a Decade of Reform.** *Public Adm Rev.* 2006; **66**(1): 77–89.  
[Publisher Full Text](#)
5. Moynihan DP: **Explaining the Implementation of Performance Management Reforms.** *The Dynamics of Performance Management.* Washington, DC: Georgetown University Press; 2008.  
[Reference Source](#)
6. Moynihan DP: **Goal-based learning and the future of performance management.** *Public Adm Rev.* 2005; **65**(2): 203–16.  
[Publisher Full Text](#)
7. Smith PC: **Measuring outcome in the public sector.** Taylor & Francis; 1996.  
[Reference Source](#)
8. Ingraham PW: **Performance: Promises to keep and miles to go.** *Public Adm Rev.* 2005; **65**(4): 390–5.  
[Publisher Full Text](#)
9. Moynihan DP, Ingraham PW: **Look for the Silver Lining: When Performance-Based Accountability Systems Work.** *J Public Adm Res Theory.* 2003; **13**(4): 469–90.  
[Publisher Full Text](#)
10. Bejerot E, Hasselblad H: **Forms of intervention in public sector organizations: Generic traits in public sector reforms.** *Organ Stud.* 2013; **34**(9): 1357–80.  
[Reference Source](#)
11. Musgrove P: **Rewards for good performance or results: A short glossary.** Washington, DC: The World Bank; 2011.  
[Reference Source](#)
12. Wang X: **Assessing administrative accountability results from a national survey.** *Am Rev Public Adm.* 2002; **32**(3): 350–70.  
[Publisher Full Text](#)
13. Streib GD, Poister TH: **Assessing the validity, legitimacy, and functionality of performance measurement systems in municipal governments.** *Am Rev Public Adm.* 1999; **29**(2): 107–23.  
[Publisher Full Text](#)
14. Grossman SJ, Hart OD: **An analysis of the principal-agent problem.** *Econometrica.* 1983; **51**(1): 7–45.  
[Publisher Full Text](#)
15. Jensen MC, Meckling WH: **Theory of the firm: Managerial behavior, agency costs and ownership structure.** *J financ econ.* 1976; **3**(4): 305–60.  
[Publisher Full Text](#)
16. Eisenhardt KM: **Agency theory: An assessment and review.** *Acad Manage Rev.* 1989; **14**(1): 57–74.  
[Publisher Full Text](#)
17. Elster J, editor: **Rational choice.** New York: NYU Press; 1986.  
[Reference Source](#)
18. Monroe KR, Maher KH: **Psychology and rational actor theory.** *Polit Psychol.* 1995; **16**(1): 1–21.  
[Publisher Full Text](#)
19. Cuevas-Rodríguez G, Gomez-Mejia LR, Wiseman RM: **Has agency theory run its course?: Making the theory more flexible to inform the management of reward systems.** *Corp Gov.* 2012; **20**(6): 526–46.  
[Publisher Full Text](#)
20. Paul E, Dramé ML, Kashala JP, *et al.*: **Performance-Based Financing to Strengthen the Health System in Benin: Challenging the Mainstream Approach.** *Int J Health Policy Manag.* 2018; **7**(1): 35–47.  
[Publisher Full Text](#)
21. Paul E, Renmans D: **Performance-based financing in the health sector in low- and middle-income countries: Is there anything whereof it may be said, see, this is new?** *Int J Health Policy Manag.* 2017.  
[PubMed Abstract](#) | [Publisher Full Text](#)
22. Blacklock C, MacPepple E, Kunutsor S, *et al.*: **Paying for performance to improve the delivery and uptake of family planning in low and middle income countries: A systematic review.** *Stud Fam Plann.* 2016; **47**(4): 309–24.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Das A, Gopalan SS, Chandramohan D: **Effect of pay for performance to improve quality of maternal and child care in low- and middle-income countries: a systematic review.** *BMC Public Health.* 2016; **16**(1): 321.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
24. Fox S, Witter S, Wylde E, *et al.*: **Paying health workers for performance in a fragmented, fragile state: reflections from Katanga Province, Democratic Republic of Congo.** *Health Policy Plan.* 2014; **29**(1): 96–105.  
[PubMed Abstract](#) | [Publisher Full Text](#)
25. Fretheim A, Witter S, Lindahl A, *et al.*: **Performance-based financing in low- and middle-income countries: still more questions than answers.** *Bull World Health Organ.* 2012; **90**(8): 559–559A.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
26. Leonard KL, Masatu MC: **Changing health care provider performance through measurement.** *Soc Sci Med.* 2017; **181**: 54–65.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
27. Miller G, Babiary KS: **Pay-for-performance incentives in low- and middle-income country health programs.** Cambridge, MA: National Bureau of Economic Research, 2013; Contract No. W18932.  
[Publisher Full Text](#)
28. Montagu D, Yamey G: **Pay-for-performance and the Millennium Development Goals.** *Lancet.* 2011; **377**(9775): 1383–5.  
[PubMed Abstract](#) | [Publisher Full Text](#)
29. Peabody JW, Shimkhada R, Quimbo S, *et al.*: **The impact of performance incentives on child health outcomes: results from a cluster randomized controlled trial in the Philippines.** *Health Policy Plan.* 2014; **29**(5): 615–21.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. Renmans D, Holvoet N, Criel B, *et al.*: **Performance-based financing: the same is different.** *Health Policy Plan.* 2017; **32**(6): 860–8.  
[PubMed Abstract](#) | [Publisher Full Text](#)
31. Witter S, Fretheim A, Kessy FL, *et al.*: **Paying for performance to improve the delivery of health interventions in low- and middle-income countries.** *Cochrane Database Syst Rev.* 2012; **(2)**: CD007899.  
[PubMed Abstract](#) | [Publisher Full Text](#)
32. Best A, Greenhalgh T, Lewis S, *et al.*: **Large-System Transformation in Health Care: A Realist Review.** *Milbank Q.* 2012; **90**(3): 421–56.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
33. MacFarlane A, Barton-Sweeney C, Woodard F, *et al.*: **Achieving and sustaining profound institutional change in healthcare: Case study using neo-institutional theory.** *Soc Sci Med.* 2013; **80**: 10–8.  
[PubMed Abstract](#) | [Publisher Full Text](#)
34. Greenhalgh T, Humphrey C, Hughes J, *et al.*: **How Do You modernize a health service? A realist evaluation of whole-scale transformation in London.** *Milbank Q.* 2009; **87**(2): 391–416.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
35. Witter S, Toonen J, Meessen B, *et al.*: **Performance-based financing as a health system reform: mapping the key dimensions for monitoring and evaluation.** *BMC Health Serv Res.* 2013; **13**(1): 367.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
36. Moynihan DP, Landuyt N: **How do public organizations learn? Bridging cultural and structural perspectives.** *Public Adm Rev.* 2009; **69**(6): 1097–105.  
[Publisher Full Text](#)
37. Moynihan DP, Pandey SK: **The big question for performance management: Why do managers use performance information?** *J Public Adm Res Theory.* 2010; **20**(4): 849–66.  
[Publisher Full Text](#)
38. Mark MM, Henry GT: **The Mechanisms and Outcomes of Evaluation Influence.** *Evaluation.* 2004; **10**(1): 35–57.  
[Publisher Full Text](#)
39. Henry GT, Mark MM: **Beyond use: Understanding evaluation's influence on attitudes and actions.** *Am J Eval.* 2003; **24**(3): 293–314.  
[Publisher Full Text](#)
40. Weiss CH, Murphy-Graham E, Birkeland S: **An Alternate Route to Policy Influence: How Evaluations Affect D.A.R.E.** *Am J Eval.* 2005; **26**(1): 12–30.  
[Publisher Full Text](#)
41. Díaz-Puente JM, Montero AC, de los Ríos Carmenado I: **Empowering communities through evaluation: some lessons from rural Spain.** *Community Dev J.* 2009; **44**(1): 53–67.  
[Publisher Full Text](#)
42. Jacob S, Ouvrard L, Bélanger JF: **Participatory evaluation and process use within a social aid organization for at-risk families and youth.** *Eval Program Plann.* 2011; **34**(2): 113–23.  
[PubMed Abstract](#) | [Publisher Full Text](#)
43. Rissi C, Sager F: **Types of knowledge utilization of regulatory impact assessments: Evidence from Swiss policymaking.** *Regulation & Governance.* 2013; **7**(3): 348–64.  
[Publisher Full Text](#)
44. Ledermann S: **Exploring the Necessary Conditions for Evaluation Use in Program Change.** *Am J Eval.* 2012; **33**(2): 159–78.  
[Publisher Full Text](#)
45. IADB: **Operating Model. Salud Mesoamerica 2015: Results Based Funding.** Washington, DC: Inter-American Development Bank; 2017; [cited 2017 Nov. 10, 2017].  
[Reference Source](#)



46. Colson KE, Potter A, Conde-Glez C, *et al.*: **Use of a commercial ELISA for the detection of measles-specific immunoglobulin G (IgG) in dried blood spots collected from children living in low-resource settings.** *J Med Virol.* 2015; 87(9): 1491–9.  
[PubMed Abstract](#) | [Publisher Full Text](#)
47. Global-Health-Workforce-Alliance: **Mid-level health workers for delivery of essential health services - A global systematic review and country experiences.** Geneva: WHO - Global Health Workforce Alliance; 2012.  
[Reference Source](#)
48. Vellez M: **Contracting-out Primary Health Care Services using Performance-Based Payments: An evaluation of the Honduras' Experience.** Rome: University of Rome II Tor Vergata; 2015.  
[Publisher Full Text](#)
49. Battye F: **Payment by Results in the UK: Progress to date and future directions for evaluation.** *Evaluation.* 2015; 21(2): 189–203.  
[Publisher Full Text](#)
50. Meessen B, Soucat A, Sekabaraga C: **Performance-based financing: just a donor fad or a catalyst towards comprehensive health-care reform?** *Bull World Health Organ.* 2011; 89(2): 153–6.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
51. Pawson R: **Evidence-based policy: A realist perspective.** Thousand Oaks, CA: Sage Publications; 2006.  
[Reference Source](#)
52. Pawson R, Tilley N: **Realistic evaluation.** Sage; 1997.  
[Reference Source](#)
53. Marchal B, van Belle S, van Olmen J, *et al.*: **Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research.** *Evaluation.* 2012; 18(2): 192–212.  
[Publisher Full Text](#)
54. Goicolea I, Vives-Cases C, San Sebastian M, *et al.*: **How do primary health care teams learn to integrate intimate partner violence (IPV) management? A realist evaluation protocol.** *Implement Sci.* 2013; 8(1): 36.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
55. Prashanth NS, Marchal B, Hoeree T, *et al.*: **How does capacity building of health managers work? A realist evaluation study protocol.** *BMJ Open.* 2012; 2(2): e000882.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
56. Van Belle SB, Marchal B, Dubourg D, *et al.*: **How to develop a theory-driven evaluation design? Lessons learned from an adolescent sexual and reproductive health programme in West Africa.** *BMC Public Health.* 2010; 10(1): 741.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
57. Vareilles G, Pommier J, Kane S, *et al.*: **Understanding the motivation and performance of community health volunteers involved in the delivery of health programmes in Kampala, Uganda: a realist evaluation protocol.** *BMJ Open.* 2015; 5(1): e006752.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
58. Deci EL, Ryan RM: **Intrinsic motivation and self-determination in human behavior.** New York: Plenum; 1985.  
[Publisher Full Text](#)
59. Gagné M, Deci EL: **Self-determination theory and work motivation.** *Journal of Organizational Behavior.* 2005; 26(4): 331–62.  
[Publisher Full Text](#)
60. Bandura A: **Self-efficacy: toward a unifying theory of behavioral change.** *Psychol Rev.* 1977; 84(2): 191–215.  
[PubMed Abstract](#) | [Publisher Full Text](#)
61. Latham GP, Borgogni L, Pettita L: **Goal Setting and Performance Management in the Public Sector.** *International Public Management Journal.* 2008; 11(4): 385–403.  
[Publisher Full Text](#)
62. Locke EA, Latham GP: **Building a practically useful theory of goal setting and task motivation. A 35-year odyssey.** *Am Psychol.* 2002; 57(9): 705–17.  
[PubMed Abstract](#) | [Publisher Full Text](#)
63. Moynihan DP: **Goal-Based Learning and the Future of Performance Management.** *Public Adm Rev.* 2005; 65(2): 203–16.  
[Publisher Full Text](#)
64. Rogers EM: **Diffusion of Innovations.** Fifth ed. New York: Free Press; 2003.  
[Reference Source](#)
65. Valente TW: **Network interventions.** *Science.* 2012; 337(6090): 49–53.  
[PubMed Abstract](#) | [Publisher Full Text](#)
66. Weyland K: **Theories of Policy Diffusion Lessons from Latin American Pension Reform.** *World Polit.* 2005; 57(2): 269–95.  
[Publisher Full Text](#)
67. Shiffman J: **Generating political priority for public health causes in developing countries: Implications from a study on maternal mortality.** CGD; 2007.  
[Reference Source](#)
68. Shiffman J: **Generating Political Priority for Public Health Causes in Developing Countries: Implications from a Study on Child Mortality.** Center for Global Development Brief Washington, DC: Center for Global Development, May Aid Policies, ed Effectiveness and Quality Department The Hague: Ministry of Foreign Affairs. 2005.
69. Shiffman J: **Issue attention in global health: the case of newborn survival.** *Lancet.* 2010; 375(9730): 2045–9.  
[PubMed Abstract](#) | [Publisher Full Text](#)
70. Greenhalgh T, Robert G, Bate P, *et al.*: **How to spread good ideas. A systematic review of the literature on diffusion, dissemination and sustainability of innovations in health service delivery and organisation.** London: University College; 2004.  
[Reference Source](#)
71. Greenhalgh T, Robert G, MacFarlane F, *et al.*: **Diffusion of Innovations in Health Service Organisations: A Systematic Literature Review.** Malden, MA: Blackwell Publishing; 2005; 581–629.  
[Publisher Full Text](#)
72. Yin RK: **Case study research: design and methods.** Thousand Oaks, Calif.: Sage Publications; 2003.  
[Reference Source](#)
73. Creswell JW, Clark VLP: **Designing and conducting mixed methods research.** Thousand Oaks, CA US: Sage Publications, Inc; 2007; xviii: 275–xviii.  
[Reference Source](#)
74. Bradley EH, Curry LA, Devers KJ: **Qualitative data analysis for health services research: developing taxonomy, themes, and theory.** *Health Serv Res.* 2007; 42(4): 1758–72.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
75. Fram SM: **The Constant Comparative Analysis Method Outside of Grounded Theory.** *Qualitative Report.* 2013; 18(1): 1–25.  
[Reference Source](#)
76. Greenhalgh T, Wong G, Jagosh J, *et al.*: **Protocol—the RAMESES II study: developing guidance and reporting standards for realist evaluation.** *BMJ Open.* 2015; 5(8): e008567.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
77. Wong G, Westhorp G, Manzano A, *et al.*: **RAMESES II reporting standards for realist evaluations.** *BMC Med.* 2016; 14(1): 96.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
78. Finlay L: **Negotiating the swamp: the opportunity and challenge of reflexivity in research practice.** *Qual Res.* 2002; 2(2): 209–30.  
[Publisher Full Text](#)
79. Snijlsteit B, Bhatia R, Rankin K, *et al.*: **3ie evidence gap maps: a starting point for strategic evidence production and use.** New Delhi: International Initiative for Impact Evaluation (3ie); Contract No.: Working Paper 28. 2017.  
[Reference Source](#)
80. Snijlsteit B, Vojtkova M, Bhavsar A, *et al.*: **Evidence gap maps—a tool for promoting evidence-informed policy and prioritizing future research.** Washington DC: The World Bank; 2013.  
[Reference Source](#)
81. Wong G, Westhorp G, Pawson R, *et al.*: **Realist synthesis: RAMESES training materials.** London, UK: National Institute for Health Research (NIHR) and Health Services Delivery Research (HSDR); 2013.
82. Rycroft-Malone J, McCormack B, Hutchinson AM, *et al.*: **Realist synthesis: illustrating the method for implementation research.** *Implement Sci.* 2012; 7: 33.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
83. Astbury B, Leeuw FL: **Unpacking black boxes: mechanisms and theory building in evaluation.** *Am J Eval.* 2010; 31(3): 363–81.  
[Publisher Full Text](#)
84. Hedström P, Ylikoski P: **Analytical sociology and rational-choice theory.** In: Manzo G, editor. *Analytical Sociology: Actions and Networks.* John Wiley & Sons; 2014; 57.  
[Publisher Full Text](#)
85. Demeulenaere P, editor: **Analytical Sociology and Social Mechanisms.** Cambridge, UK: Cambridge University Press; 2011.  
[Publisher Full Text](#)
86. Hedström P, Wennberg K: **Causal mechanisms in organization and innovation studies.** *Innovation.* 2017; 19(1): 91–102.  
[Publisher Full Text](#)



## Open Peer Review

Current Peer Review Status:   

Version 1

Reviewer Report 05 March 2018

<https://doi.org/10.21956/gatesopenres.13842.r26179>

© 2018 Hirschhorn L. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Lisa R. Hirschhorn** 

Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

The authors have developed an in depth and well written description of the rationale behind and approaches to the protocol for a study using a realist evaluation approach to do an intern evaluation of AMI, a large multicounty accountability-driven intervention. The approach will complement the plan program evaluation being completed by IHME which is largely looking at explicit program defined outcomes. The description off SMI particularly for readers not as familiar with the structure is very helpful.

The authors are clearly fluent in Realist Evaluation and familiar with many of the underlying theories which they use. However there is a lack of clarity of what the main focus is of this manuscript is describing and the use of the term "study" is often confusing as referring to different scopes of work.

For example:

Abstract:

The initial sentence which reads " This **study** presents the protocol for a **study** that uses a realist evaluation approach to develop a preliminary program theory that hypothesizes the interactions between context, interventions and the mechanisms that trigger outcomes. The program theory was completed through a scoping review of relevant empirical, peer-reviewed and grey literature; a sense-making workshop with program stakeholders; and content analysis of key SMI documents." And then goes onto to say "This **study**".

In the text, the reviewer was still confused which study was being described (the development, the testing, et he evaluation leading to results including a refined program theory) and clarity would be helpful, including that the protocol describes work already done (development of the preliminary program theory) as well as how it will be applied in the future.

In framing the manuscript in the text, later they then state " This study addresses two research questions: "(1) What are the effects of using supply-side financial incentives on the performance

of the primary care systems in Honduras and El Salvador? How are those effects produced? Under what contextual factors are these effects produced in each country? And, (2) What are the effects of continuous external verification of performance in the two countries under study? How are those effects produced? Under what contextual factors are these effects produced in each country?

While I assume that this use of the term: study" refers to the realist evaluation rather than the development of the program theory. For example in the section " Study design" it states ". In this step the preliminary program theory will be tested, further developed, and validated or rejected."

Given the critical importance of the qualitative data to be collected through interviews, a bit more detail in how the interviewees will be sampled (site, individual, area in the respective countries)

In their challenges part, it would be helpful to understand a bit more the limitations imposed by the 2 countries chosen from SMI for this study, and what characteristics differ from other SMI countries not chosen for this evaluation

Minor:

On page 8 in describing the program theory, I am curious that inputs are not explicitly called out as needed (and related to context) and that equity and effectiveness are also not explicit in the theory.

Given the design of SMI and the underlying approach of Realist Evaluation, I was curious if the researchers had considered including community interviewer and or patients as critical to the success (and acceptability) of the intervention.

Are they also planning to assess fidelity to the planned implementation (and adaptations implemented locally or at a national level) which could change the outcomes and be related to or change the mechanisms (as well as inform potential future adaptations.

**Is the rationale for, and objectives of, the study clearly described?**

Yes

**Is the study design appropriate for the research question?**

Yes

**Are sufficient details of the methods provided to allow replication by others?**

Yes

**Are the datasets clearly presented in a useable and accessible format?**

Not applicable

**Competing Interests:** No competing interests were disclosed.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 26 Sep 2018

**Wolfgang Munar**, George Washington University, Washington, USA

Lisa- Thanks a lot for your comments. The authors have reviewed them all. See below the specific actions we have taken.

**Comment** - There is a lack of clarity of what the main focus is of this manuscript is describing and the use of the term “study” is often confusing as referring to different scopes of work. In the text, the reviewer was still confused which study was being described (the development, the testing, the evaluation leading to results including a refined program theory) and clarity would be helpful, including that the protocol describes work already done (development of the preliminary program theory) as well as how it will be applied in the future.

### Response

We very much appreciate the reviewer’s observations about the temporal relationships among the various components of this large scale, multi-year, multi-phase evaluation. We have carefully reviewed the manuscript and made edits to clarify tense accordingly. Regarding the development of the program theory, we explicitly state that the theory has been developed before data collection (page 7- Preliminary program theory section), as per the standards of realist evaluation practice (Wong, Westhorp et al. 2016). We also describe briefly how the program theory will be applied and assessed in the subsequent phase of work.

In framing the manuscript in the text, later they then state “This study addresses two research questions: “(1) What are the effects of using supply-side financial incentives on the performance of the primary care systems in Honduras and El Salvador? How are those effects produced? Under what contextual factors are these effects produced in each country? And, (2) What are the effects of continuous external verification of performance in the two countries under study? How are those effects produced? Under what contextual factors are these effects produced in each country? While I assume that this use of the term: study” refers to the realist evaluation rather than the development of the program theory. For example, in the section “Study design” it states “. In this step the preliminary program theory will be tested, further developed, and validated or rejected.”

**Response:** The reviewer is correct. In the instance noted, we use the term ‘study’ to refer to

the full, multi-method, multisite realist evaluation of SMI. The program theory is preliminary work, as described on page 7, (see the section introducing the preliminary program theory).

**Comment** - Given the critical importance of the qualitative data to be collected through interviews, a bit more detail in how the interviewees will be sampled (site, individual, area in the respective countries)

### Response

The steps and sequence of the realist evaluation has been clarified, and further details about the data collection process have been added. See Methods section (pages 6-10).

**Comment** - In their challenges part, it would be helpful to understand a bit more the limitations imposed by the 2 countries chosen from SMI for this study, and what characteristics differ from other SMI countries not chosen for this evaluation

### Response

We have clarified the rationale for choosing the two high-performing countries in the Methods section and expanded the ensuing limitations in the Discussion section (pages 12-13)

**Comment** - On page 8 in describing the program theory, I am curious that inputs are not explicitly called out as needed (and related to context) and that equity and effectiveness are also not explicit in the theory.

### Response

Program theory in realist evaluation is not based on conventional logical models that use input-process-output-outcome configurations. PT in realist evaluation are not equivalent to theories of change, either. PT as used and detailed in the updated version of the protocol refer to context-mechanism-outcome configurations that are informed by existing empirical evidence, social science theories, and input from program stakeholders. We also agree with the reviewer's comments about effectiveness and equity. These aspects are detailed in SMI's original theory of change and in (now) tables 1 and 2. It is important to note, however, that the review of the literature indicates that such long-term or distal outcomes are unlikely to be measurable at the mid-term stage in which the evaluation will take place.

**Comment** - Given the design of SMI and the underlying approach of Realist Evaluation, I was curious if the researchers had considered including community interviewer and or patients as critical to the success (and acceptability) of the intervention.

### Response

The suggested approach would have been ideal. However, operational constraints that are now described in the Discussion section (pages 12-13) made such design options not feasible for this first evaluation. We agree with the reviewer that the inclusion of a demand-side perspective is highly advisable for future iterations of SMI evaluation.

**Comment** - Are they also planning to assess fidelity to the planned implementation (and adaptations implemented locally or at a national level) which could change the outcomes

and be related to or change the mechanisms (as well as inform potential future adaptations).

### Response

The realist evaluation will not assess fidelity to planned implementation, but will identify and explore country adaptations. These aspects of flexibility in implementation and country adaptation are addressed in the Methods section (page 6).

### Reference

Wong, G., G. Westhorp, A. Manzano, J. Greenhalgh, J. Jagosh and T. Greenhalgh (2016). "RAMESES II reporting standards for realist evaluations." BMC Medicine **14**.

**Competing Interests:** No competing interests were disclosed.

Reviewer Report 30 January 2018

<https://doi.org/10.21956/gatesopenres.13842.r26182>

© 2018 Dossou J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



### Jean-Paul Dossou

<sup>1</sup> Centre de Recherche en Reproduction Humaine et en Démographie, CNHU/HKM, Cotonou, Benin

<sup>2</sup> Institute of Tropical Medicine of Antwerp, Antwerp, Belgium

This is a brilliant manuscript, among the best I have ever reviewed. The subject is relevant and this paper will serve several academic and scientific purposes.

The following comments and questions may help in improving some minor points.

1. In which regards are El Salvador and Honduras contrasting cases? Can authors provide a brief comparison table showing in which dimensions those countries are considered contrasting cases?
2. Figure 2: Improve the display of the four squares of the "scaling-up of interventions" box.
3. Study design/1st paragraph  
Authors reported the following "we define each country's primary care system as the unit of analysis". Can authors provide an operation definition/conceptual framework of "country's primary care system" within this protocol?
4. Data analysis/4th paragraph  
"preliminary program theory and the causal patterns identified." not "preliminary program thyeory and the causal patterns identified."



# 5. Data analysis/1st paragraph

We suggest to authors to include "actors " in the "context, intervention, mechanism, and outcome " structure to have "intervention, context, actor, mechanism, and outcome (ICAMO) " like here (<https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-017-4322-8#CR42>). Authors may also broaden the CMO configuration to consider the ICAMO configuration that may improve quality in the analysis and make a better and more explicit use of the role of actors in the analysis.

## Is the rationale for, and objectives of, the study clearly described?

Yes

## Is the study design appropriate for the research question?

Yes

## Are sufficient details of the methods provided to allow replication by others?

Yes

## Are the datasets clearly presented in a useable and accessible format?

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Health policy and system research

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 30 Jan 2018

**Wolfgang Munar**, George Washington University, USA

Dr. Dossou: Les auteurs apprécient vos commentaires. Merci beaucoup.

We will consider them all, while editing the paper.

**Competing Interests:** No competing interests were disclosed.

Author Response 26 Sep 2018

**Wolfgang Munar**, George Washington University, Washington, USA

Jean-Paul: Thanks for your comments. The authors have revised the study protocol based on yours and the excellent comments from other reviewers. See below 2 specific comments in response to your suggestions.

**Comment** - In which regards are El Salvador and Honduras contrasting cases? Can authors provide a brief comparison table showing in which dimensions those countries are considered contrasting cases?

**Response**

The study setting section (see page 5) describes the major distinctions in institutional context between the two countries.

**Comment** - We suggest to authors to include "actors" in the "context, intervention, mechanism, and outcome" structure to have "intervention, context, actor, mechanism, and outcome (ICAMO)" like here (<https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-017-4322-8#CR42>). Authors may also broaden the CMO configuration to consider the ICAMO configuration that may improve quality in the analysis and make a better and more explicit use of the role of actors in the analysis.

**Response**

We appreciate the recommendation. However, after consideration we decided to follow the standards in reporting realist evaluations developed in 2016 (Wong, Westhorp et al. 2016) which recommend using CMO configurations.

**References**

Wong, G., G. Westhorp, A. Manzano, J. Greenhalgh, J. Jagosh and T. Greenhalgh (2016). "RAMESES II reporting standards for realist evaluations." BMC Medicine **14**.

**Competing Interests:** No competing interests were disclosed.

Reviewer Report 12 January 2018

<https://doi.org/10.21956/gatesopenres.13842.r26180>

© 2018 Kress D. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Daniel H. Kress**

RTI International, Seattle, WA, USA

I recommend publication with minor revisions.

As this is a study protocol as opposed to the actual study, there are no datasets at this time so the answer to the question, "Are the datasets clearly presented in a useable and accessible format" can only be partly or actually NA since the data will be collected using the study protocol that is proposed for publication and eventually that will be carried out to assess the impact of SMI.

Overall, I find this to be a thorough and carefully thought out study protocol that will provide important insights into how the results produced by SMI were actually created. As such, this study protocol will shed important insights into how a large, complex intervention across multiple

countries and over time produced the quite astounding results that marked the success of SMI. Even as we have seen the positive results from the regular evaluations and can easily see the quite significant improvements countries that are part of this initiative have registered, important questions as to what factors actually drove the impact seen remain only partially answered. This study will shed important light on these questions.

I only have a few minor quibbles regarding the article.

1. The authors use PBF and RBF almost interchangeably and sometimes use both terms. I think it might be less confusing to the reader to define terms up front and then use one term.
2. Page 3, paragraph 8, says that studies on the effects of RBF on large scale system reforms are largely absent. Later on the authors cite a systematic review. In fact, there have been a number of systematic reviews of RBF programs beyond the one cited. For example, Andy Oxman has several papers that review (critically) the experience with RBF. Miller and Singer (2013) is another.
3. I also think that in the area of RBF, it's important to not focus only on LMIC experience as RBF is an instrument that has been used and is being used extensively. The Quality and Outcomes Framework (QOF) in the UK NHS is an example. Peter Smith has a number of papers that reviews that experience and Cheryl Cashin and Peter Smith have a paper on how RBF links to the larger issue of Strategic Purchasing.
4. Perhaps my strongest comment is on page 7, paragraph 6, regarding the program theory section. I think it's quite possible to formulate a hypothesis that SMI was not primarily a classic extrinsic financial incentive program but possibly much more an extrinsic non pecuniary program where the rewards were doing well amongst your peers. When you look at the incentive rewards, its difficult to see how such relatively small financial rewards could incent behavior. The counterpoint to this argument might be that the funding provided by the SMI donors was flexible and in these heath systems flexible funding is often rare and highly prized but that too is an issue deserving of further investigation. However, if the funding is small and relatively insignificant, the question is then what drove the behavior and actions taken. A factor worthy of investigation is the SMI approach of engaging multiple countries in a form of joint competition. Ministers of Health were all engaged on SMI and there is some anecdotal evidence that the approach of having them compete together, each trying to attain the targets they set for their own country, created a form of competition or at least a common forum where not performing well would be seen as a distinct negative outcome, thereby conferring strong incentives for them to perform well or endeavor to make sure their health system performs strongly. This kinds of peer effects are known to be powerful in behavioral economics and so we should look for them in this study as well.

**Is the rationale for, and objectives of, the study clearly described?**

Yes

**Is the study design appropriate for the research question?**

Yes

**Are sufficient details of the methods provided to allow replication by others?**

Yes

**Are the datasets clearly presented in a useable and accessible format?**

Partly

**Competing Interests:** I was Deputy Director at the Bill and Melinda Gates Foundation during the time that the SMI program was designed and implemented. I also was responsible for the SMI program for two years. I know and used to work with the lead author of this article when we were both employed by the Bill and Melinda Gates Foundation.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 30 Jan 2018

**Wolfgang Munar**, George Washington University, USA

Dear Dan,

The entire team read your comments. We appreciate them enormously and will tackle them in our upcoming edited and final version.

Thanks a lot,

Wolfgang on behalf of the team.

**Competing Interests:** No competing interests were disclosed.

Author Response 26 Sep 2018

**Wolfgang Munar**, George Washington University, Washington, USA

Dan- We have followed your comments to make a revision of the study protocol. Here are some specific actions we have taken.

**Comment** - The authors use PBF and RBF almost interchangeably and sometimes use both terms. I think it might be less confusing to the reader to define terms up front and then use one term. Page 3, paragraph 8, says that studies on the effects of RBF on large scale system reforms are largely absent. Later on, the authors cite a systematic review. In fact, there have been a number of systematic reviews of RBF programs beyond the one cited. For example, Andy Oxman has several papers that review (critically) the experience with RBF. Miller and Singer (2013) is another. I also think that in the area of RBF, it's important to not focus only on LMIC experience as RBF is an instrument that has been used and is being used extensively. The Quality and Outcomes Framework (QOF) in the UK NHS is an example. Peter Smith has a number of papers that reviews that experience and Cheryl Cashin and Peter Smith have a paper on how RBF links to the larger issue of Strategic Purchasing.

**Response**

These comments made the team reinforce and rewrite the theoretical basis for the study protocol. We also explicitly linked the frameworks in the updated version to the literature on performance measurement and management, which owes a lot to the British experiences mentioned by the reviewer. Performance-based financing, pay-for-performance, and results-based financing are now subsumed under the category of “financial arrangements” as per the typology of interventions developed by the Cochrane Collaboration Effective Practice and Organization of Care (EPOC). See Introduction section, pages 2-4; and table 1.

**Comment** - Perhaps my strongest comment is on page 7, paragraph 6, regarding the program theory section. I think it's quite possible to formulate a hypothesis that SMI was not primarily a classic extrinsic financial incentive program but possibly much more an extrinsic non pecuniary program where the rewards were doing well amongst your peers. When you look at the incentive rewards, its difficult to see how such relatively small financial rewards could incent behavior. The counterpoint to this argument might be that the funding provided by the SMI donors was flexible and in these health systems flexible funding is often rare and highly prized but that too is an issue deserving of further investigation. However, if the funding is small and relatively insignificant, the question is then what drove the behavior and actions taken. A factor worthy of investigation is the SMI approach of engaging multiple countries in a form of joint competition. Ministers of Health were all engaged on SMI and there is some anecdotal evidence that the approach of having them compete together, each trying to attain the targets they set for their own country, created a form of competition or at least a common forum where not performing well would be seen as a distinct negative outcome, thereby conferring strong incentives for them to perform well or endeavor to make sure their health system performs strongly. This kinds of peer effects are known to be powerful in behavioral economics and so we should look for them in this study as well.

**Response**

We agree with the comment. However, this study is not funded to conduct a contrasting case study design at the policy level of all participating countries. However, if the hypothesized supra-national mechanism suggested by the reviewer were to exist, it would be reflected in our findings. The latter is hypothetically plausible given that we will be looking to explore the effects that global and regional (i.e., Mesoamerican) issue-specific agendas had on the decision by high-level policy makers to join SMI.

**Competing Interests:** No competing interests were disclosed.

---

## Comments on this article

Version 1

Reader Comment 20 Feb 2018



**Jennifer Nelson**, Interamerican Development Bank, Salud Mesoamerica, USA

In general, we find this study protocol to be innovative and well designed, and its research will contribute to an important research gap.

We felt that in the final version of the paper, the following should be addressed:

1) Clear definition of what the authors mean with certain terms in the context of this paper including: system performance, government performance, performance management, performance improvement, performance based results, reform, RBF, and PBF. In the context of SMI, there has been much debate on what we are measuring in terms of system performance. For example, does system performance refer to the health systems ability to meet targets, accelerate change, or sustain changes? Although the definition of performance improvement is evolving, authors should state how they are defining “system performance” and “government performance” in the context of this research paper. Regarding RBF and PBF, the paper provides a brief description of these two terms, but they are used interchangeably.

2) Characterization of SMI: we have been in internal discussions regarding what is the correct characterization and categorization of SMI in the RBF/PFB terminology. We feel that RBF “plus” is the best description, given that the three main levers used in implementation include: 1) high level financial incentive; 2) external evaluation; and 3) tailored technical assistance. The preliminary program theory focuses on high-level incentives and continuous external verification of performance, however it is important to highlight the importance of technical assistance, in addition to other factors, that have been shown to be important in other research about SMI including regionality, technical assistance, and reflective learning environment (El Bcheraoui et al., 2017). To this point, we feel it is extremely important to point out that the scope of this research focuses on only a subset of the critical pathways of change of SMI, and should not lead readers to assume that these points are only important factors in SMI. We recommend that the authors explicitly state this in the paper, including why/how the factors included were selected, and that they are not the only interventions and mechanisms included in the SMI ToC. These points should be strengthened both under study setting, methodological approach, and in Figure 2. Preliminary program theory.

We have the following specific comments for the authors:

1. Please include in paragraph 1 under *Study Setting* that reimbursed funds are non-earmarked funds for governments to use within the health sector, and are the financial incentive in the SMI model.
2. Please correct 3<sup>rd</sup> paragraph under *Study Setting*: the 1<sup>st</sup> phase of SMI focused on process and output indicators; phase 2 & 3 focus on coverage, quality and outcome indicators. Currently, paper states “During phase 2, targets were focused on outputs...”
3. Please mention in paragraph 3 under *study setting* that IHME does not just measure achievement of results included in the performance framework (10 indicators), but also measures a comparable menu of indicators called the regional performance framework. Additionally, breastfeeding is not a payment indicator due the sample size required.

***Competing Interests:*** The comments reflected here been reviewed and approved by the Salud Mesoamerica Coordination Unit. This unit manages implementation of the Initiative.

-----