



STUDY PROTOCOL

REVISED **Characterizing performance improvement in primary care systems in Mesoamerica: A realist evaluation protocol**
[version 2; peer review: 2 approved, 1 approved with reservations]

Wolfgang Munar ¹, Syed S. Wahid¹, Leslie Curry²

¹Milken Institute School of Public Health, George Washington University, Washington, DC, 20052, USA

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

V2 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. Evaluations of performance measurement and management interventions in public, primary care delivery systems of low- and middle-income countries are scarce. In such contexts, few studies to date have focused on characterizing how, why and under what contextual conditions do such complex, multifaceted arrangements lead to intended and unintended consequences for the healthcare workforce, the healthcare organizations involved, and the communities that are served.

Methods. Case-study design with purposeful outlier sampling of high-performing primary care delivery systems in El Salvador and Honduras, as part of the Salud Mesoamerica Initiative. Case study design is suitable for characterizing individual, interpersonal and collective mechanisms of change in complex adaptive systems. The protocol design includes literature review, document review, non-participant observation, and qualitative analysis of in-depth interviews. Data analysis will use inductive and deductive approaches to identify causal patterns organized as 'context-mechanism-outcome' configurations. Findings will be triangulated with existing secondary data sources collected including country-specific performance measurement data, impact, and process evaluations conducted by the Salud Mesoamerica Initiative.

Discussion. This realist evaluation protocol aims to characterize how, why and under what conditions do performance measurement and management arrangements contribute to the improvement of primary care system performance in two low-income countries.


Open Peer Review

Approval Status ✓✓?

	1	2	3
version 2 (revision) 04 Oct 2018			
version 1 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.

Keywords

El Salvador, Honduras, Primary Care Systems, Performance Measurement and Management Systems, Low- and middle-income countries, Realist Evaluation, Salud Mesoamerica Initiative

Corresponding author: Wolfgang Munar (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 2; peer review: 2 approved, 1 approved with reservations]** Gates Open Research 2018, 2:1 <https://doi.org/10.12688/gatesopenres.12782.2>

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

REVISED Amendments from Version 1

Based on the feedback received from the reviewers, the authors have introduced the following revisions to this study protocol:

1) A multi-disciplinary framework has been introduced to characterize performance measurement and management (PMM) interventions; 2) A new discussion section has been introduced describing the pros and cons of the study design selected; and, 3) The methods section, while not changed in structure, better describes the sequencing of activities.

Specifically, [Figure 2](#) has been revised, a new [Table 1](#) has been added and the previous Table 1 is now labelled [Table 2, Box 1](#) was added in response to the referee reports, and [Supplementary File 1](#) has been replaced

See referee reports

Introduction

Calls have been made for improving performance of primary care systems in low- and middle-income countries (LMICs) as a necessary condition to achieve universal health coverage in the age of the Sustainable Development Goals. High-performing primary care systems not only are the first point of contact for continuous, coordinated, comprehensive and people-centered health services¹, but also provide critical preparedness and response to global, public health threats².

There is growing interest in better understanding the ways in which various policies and programs can improve primary care health systems at scale and moving beyond the quick fixes that characterize most efforts at changing complex system change³. Large-scale, health system change is described as “coordinated, system-wide change affecting multiple organizations and care providers, with the goal of significant improvements in the efficiency of health care delivery, the quality of patient care, and population-level patient outcomes”⁴.

Organizational performance refers to the results generated by an organization, measured against its intended goals and targets. In the private sector, the concept usually refers to profits, efficiency, quality, market-share, and customer satisfaction. In public sector organizations, the definition has shifted with the evolving framings for the role of the State in the production and delivery of public services⁵. Governments’ interest has shifted from controlling inputs and compliance with standards, towards reporting quantity and quality of outputs, productivity, efficiency and, more recently, outcomes and policy impacts^{5,6}.

Performance measurement and management (PMM) systems are organizational arrangements aimed at measuring organizational processes, outputs and outcomes with the proximal aim of informing the introduction of clinical, managerial, programmatic and policy changes, and the ultimate purpose of contributing to socially valued, population-level health and equity outcomes⁷. Forty years of research on PMM systems have shown that such systems can effectively improve performance, although unintended and undesirable effects can also occur^{8–16}. While there have been applications of some types of PMM interventions to the health sector of LMICs particularly through

the use of financial incentives and pay-for-performance, research that bridges advances made in public administration and organizational science remains largely ignored in health systems research.

In order to address this fragmentation in evidence, we developed a framework that combines a PMM model originally developed by Pollitt to study organizational performance in the public sector¹⁷ and a taxonomy developed by the Cochrane Collaboration’s Effective Practice and Organization of Care (EPOC) to characterize through the use interventions and outcomes in healthcare delivery¹⁸. The former helped us define the main elements of system-wide PMM interventions, while the latter allowed us to identify PMM interventions of relevance to primary care delivery systems in LMICs.

The general PMM system framework contains the following components: (1) An institutional context in which various policies, programs and health interventions are implemented and interact with healthcare stakeholders; (2) a local, socio-economic context where primary care services are delivered; (3) one or more PMM interventions that trigger improvements (or not); (4) a performance measurement process; (5) a sense-making process that allows the transformation of raw data into performance information; (6) a process of dissemination of performance information among system actors and stakeholders with the intent of making it actionable; (7) performance information use, misuse, or non-use; (8) implementation of planned action, leading to measurable organizational improvements (or not); and, (9) the production (or not) of short-term clinical and managerial improvements; intermediate outputs and outcomes; and, distal, societal, and population-level health and equity outcomes (intended and otherwise).

The EPOC taxonomy, in turn, contains various cross-cutting interventions and organizational arrangements of relevance to primary care systems such as implementation strategies, accountability arrangements, and some examples of financial arrangements. Such interventions can induce performance improvements at the level of the workforce, facilities, patients, and populations^{16,19–23}. Furthermore, we hypothesized that PMM interventions may operate at individual (providers, managers, etc.) and/or organizational-levels (facilities, networks of care, local health systems, etc.) and can trigger outcomes across short and long timeframes (desirable as well as undesirable, adverse effects). The main types of PMM interventions and outcomes of relevance to primary care delivery systems are listed in [Table 1](#).

Primary care performance improvement in LMICs has been mostly studied to date by means of research that addresses the effective delivery of health interventions by providers and facilities. Such studies rarely address the effects of PMM interventions on the behaviors of providers and facilities charged with service delivery. We believe that the understanding of health system performance improvement processes requires research that characterizes the context, mechanisms, and processes through which various PMM interventions trigger (or not) process improvements, organizational learning, and system-wide adaptation including but not limited to the emergence of quality supply,

Table 1. PMM interventions and outcomes.

PMM System Components	Definitions and examples
INTERVENTIONS	
Implementation strategies	Refer to interventions designed to bring about changes in healthcare organization, in the behavior of healthcare professionals or in the use of health services by recipients such as in-service training, continuing education, reminders, supervision, clinical guidelines, clinical incident reporting, and continuous quality improvement, among others ^{18–21} .
Accountability arrangements	Refer to the organizational and institutional interventions used in public administration to verify and control the delivery of public services and can include, among others, the provision of audit and feedback to providers ^{16,22–26} , or the use of social accountability interventions like the public release of performance information and community monitoring ^{27–31} .
Financial arrangements	Refer to changes in how funds are collected, how services are purchased, and the use of insurance schemes as well as financial incentives or disincentives. In this evaluation, we will solely focus on financial interventions that have performance-improvement potential such as the use of rewards or incentives (financial and in-kind) and performance-based financing ^{15,32} .
OUTCOMES	
Provider and managerial outputs and outcomes	Provider and managerial outputs: Individual, provider and managerial staff effects, and exemplified by changes in workload, work morale, stress, burnout, sick leave, and staff turnover
Patient outcomes (changes in health status or on patient health behaviors)	Physical health and treatment outcomes; adherence to treatment or care plans by patients, and/or health-seeking behaviors; and unintended patient outcomes
Organizational outcomes (organization-level, within and across-facilities of care)	Quality of care process improvements, patient satisfaction, perceived quality of care, workforce retention, organizational culture, and unintended outcomes (gaming, shirking, shaming, data falsification, etc.)
Policy effects (changes in rules and regulations)	New rules, regulations, guidelines, protocols of care, etc.
Population-level outputs and outcomes (aggregate, health and equity effects accruing defined populations, including utilization of specific primary care services)	Number of antenatal care visits, institutional deliveries, etc.), coverage of services (such as the proportion of pregnant women receiving antenatal care, proportion of pregnant women delivering in facilities; coverage rate of specific vaccines), access to primary care services (for instance, waiting times), adverse health effects or harm, health equity effects, and unintended health effects, etc.
Social outcomes (non-health, social, economic, or cultural effects affecting defined populations)	Changes in community participation; non-health equity effects; non-health adverse effects or harm; and, other unintended social outcomes,

Adapted from: 1) EPOC (2015). EPOC Taxonomy, The Cochrane Collaboration; 2) EPOC- The EPOC taxonomy of health systems interventions. Resources for review author. Oslo, Norway, Norwegian Knowledge Centre for the Health Services.

patient safety, and population-level equity and health outcomes. Such research is necessary in support of recent calls for a revolution in quality health systems in global health settings³.

This evaluation protocol aims to characterize how, why and under what contextual conditions has the Salud Mesoamerica Initiative (SMI) triggered performance improvements in El Salvador and Honduras through the introduction of various types of PMM arrangements. In the next section we introduce SMI, and in subsequent sections we describe the rationale for the evaluation, the methods to be employed, and further discuss the strengths and limitations of the proposed research.

Study setting

SMI is a multi-country, large-scale PMM initiative resulting from the partnership between the governments of the eight Mesoamerican nation-states, the Bill and Melinda Gates Foundation, the Carlos Slim Foundation, the Government of Canada, and the Inter-American Development Bank (IADB). SMI is a performance-based financing program that supports participating

governments' production of population-level health and equity outcomes through arrangements that ultimately aim at improving primary care delivery for the poor at scale.

The program was sequenced as three consecutive phases of eighteen to twenty-four months each, for the achievement of progressively complex performance targets in reproductive, maternal, neonatal and child health. Phase 1 programs started in a staggered fashion in 2011 and the final stage started in 2018 and will end in 2020. Performance targets during phase 1 had an initial focus on adherence to standards of care, availability of supplies and, in general, process and output targets. During phases 2 and 3, targets prioritize outcomes such as modern contraceptive prevalence, effective coverage of antenatal care and institutional deliveries, post-partum and post-natal care coverage, and in some countries, reductions in the prevalence of anemia and gains in immunological coverage of measles vaccination^{24–26}.

At baseline, IHME collected data from 20,225 households and 479 primary care facilities in the poorest, rural municipalities

of all participating countries. Results varied significantly between and within countries, underlying differences in health system performance, availability of inputs, quality of services, and highlighting poverty-related and other disparities in health outcomes^{26,27}.

Upon joining SMI, participating governments contributed domestic funds and formally agreed with the IADB to a set of performance targets for each of the three phases. The IADB then matched domestic contributions with grant financing on a 1:1 ratio. Performance contracts between the IADB and each government provided that the former would reimburse half of the initial domestic investment, contingent on the achievement of 80% or more of the agreed-upon targets. Measurement of programmatic performance by an external, mutually trusted agency (i.e., IHME), was required to ensure accountability and credibility in results.

SMI's original theory of change (Figure 1) hypothesized that the supply-side financial incentives would target ministries of health (MOH) attention on achieving the agreed performance targets and that the latter would be reinforced by the external measurement of performance. These cycles would be further reinforced by ongoing technical support, policy dialogue, and purposeful dissemination of performance information. Such processes would, in turn, lead to progressive improvements in the availability of quality supply and enhanced, aggregate performance in the primary care delivery system. Additional causal assumptions rested on an increase in domestic pro-poor health

spending, and an expansion in the demand for high-impact health interventions among beneficiary populations.

In 2011, the partners agreed on a set of common, high-level principles such as a focus on results, independent performance measurement, and mutual accountability and transparency. 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. The implementation approaches through which the program's PMM interventions would be transmitted downstream into the delivery systems were not prescribed *a-priori* by SMI and were, instead, left to country-specific, flexible implementation arrangements.

In the two countries under study, El Salvador and Honduras, the focus on country ownership led to each government deciding how to deploy SMI's non-reimbursable resources and their own domestic financing for the achievement of program targets. El Salvador had gone through a health system reform in the late 2000s, which coincided with the beginning of SMI implementation. There, the government decided to focus its targets on the provision of universal primary care services through Community Health Teams²⁸, one of the reform's central features. Honduras, in turn, had started a large-scale contracting-out and pay-for-performance programs in the late 2000s²⁹. The government decided to leverage its experience with those PMM financial arrangements and implemented SMI in primary care systems that had already acquired experience with PMM

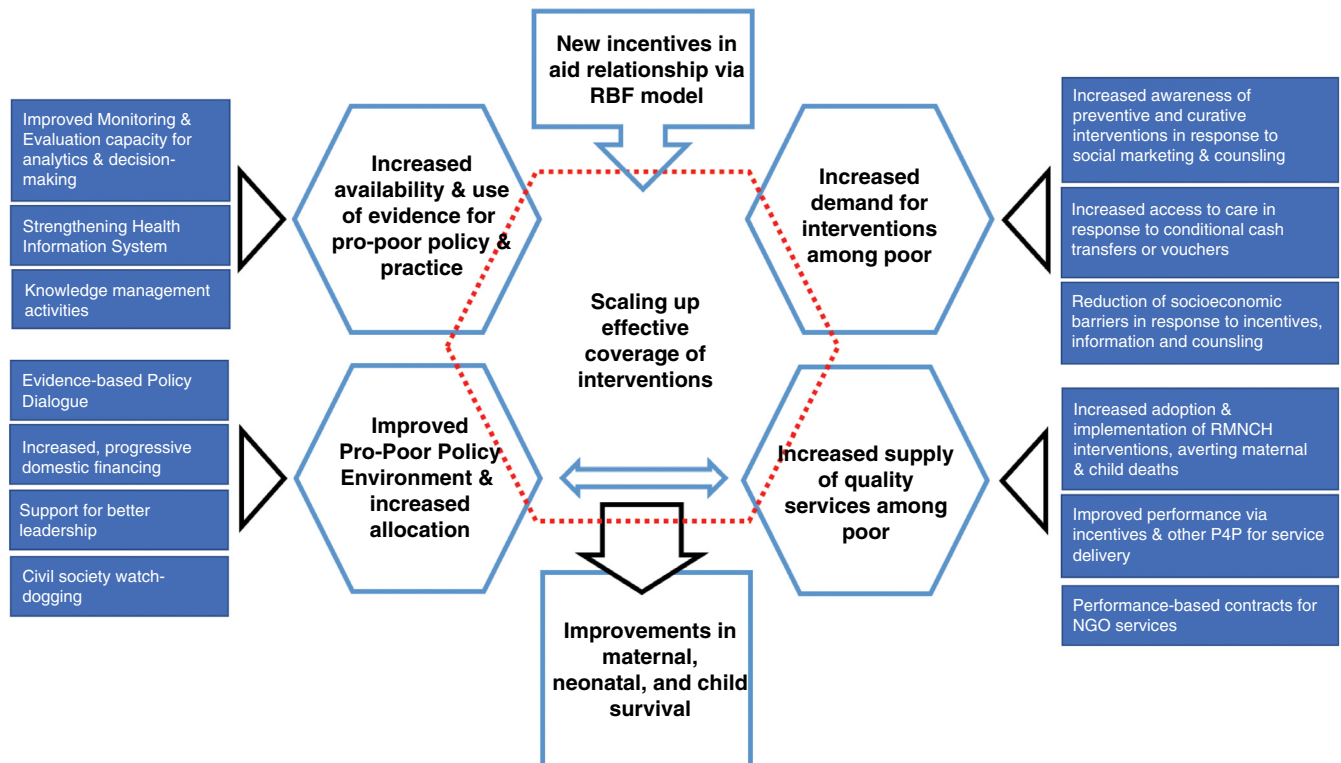


Figure 1. SMI initial theory of change.

arrangements. [Table 2](#) lists some of the targets agreed by El Salvador and Honduras.

Methods

This study protocol addresses two research questions: (1) What are the effects of using supply-side incentives on the performance of primary care systems in El Salvador and Honduras? How are those effects produced and under what contextual conditions? And, (2) What are the effects of external measurement of performance on the primary care systems of El Salvador and Honduras? How are those effects produced and under what contextual conditions?

We recognize that the evaluation of a program as complex as SMI needs to be informed by methodological approaches that go beyond the measurement of progress against agreed-upon performance targets and should also attempt to further explore the lessons that can be learned from the flexible, adaptive nature imagined by SMI and its sponsors. By providing governments with a high degree of flexibility in implementation, SMI introduced important distinctions with other global program partnerships. It explicitly attempted to increase government buy-in, and encouraged local adaptation and learning which, in turn, make concerns with fidelity in implementation less important than, for instance, characterizing the adaptations that worked or not, and why. Therefore, in this study protocol we follow the approach used in the evaluation of other whole-system transformational reforms which suggest that “program fortunes can be shaped and constrained by interactions between the program and the context”³⁰ in each participating country and, also, by the necessary adaptations and responses to dynamic and changing environmental conditions.

To address the research questions and the complex dynamics introduced by SMI’s adaptive approach, we decided to use realist evaluation. Realist evaluation is based on the premise that an evaluation needs to answer “what worked, how, in what circumstances and for whom”^{31,32}. It is a form of theory-driven program evaluation that has been used in evaluation studies and in health systems research for the evaluation of complex policies, programs and interventions in various socio-economic settings, including LMICs^{33–44}. 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. Such perspective contends that program interventions bring about change through underlying, usually hidden, causal mechanisms, and considers the role of context as indispensable in explaining causality.

The starting point in a realist evaluation is the development of a Program Theory (PT). In this study protocol, the preliminary PT was developed based on previous research identified through literature review, document review, and consultations with experts involved in SMI’s design, implementation, and evaluation. The PT will be used to inform the process of data collection and to the completion of a refined PT. The latter will provide explanations of why, how and under what

conditions do SMI interventions trigger causal mechanisms that, in turn, lead to specific outcomes, intended and otherwise.

This evaluation is an 18-month study running from May 2017 to December 2018 and executed contemporaneously with the finalization of SMI’s phase 2 in El Salvador and Honduras. The evaluation seeks to maximize diversity in institutional and policy context to increase the likelihood of identifying variations in policy and program conditions and thus characterizing the process of change generated to date by the program.

At the country-level, a case-study design with contrasting cases was selected as the primary study design. We defined each country’s primary care system as the unit of analysis. Furthering the purpose of this evaluation to understand high-performance at large-scales, we will also purposefully identify and study outlier, high-performing primary care delivery systems which “can reveal a great deal about intense manifestations of the phenomenon of interest”⁴⁵. Contrasting case approaches align well with the realist evaluation proposition that contexts can trigger to-be-identified mechanisms that, in turn, interact with program interventions and contribute to generating outcomes (or not).

Preliminary program theory

This step has already been completed. For the development of the preliminary PT, we first reviewed the literature to identify social science theories and empirical evidence that explicitly addressed PMM interventions and outcomes in public and private organizations; characterized the mechanisms of change of relevance for large-scale health system change; and, explored the scarce evidence that exists about the role of context in triggering or obstructing health system transformation. Given that one of the authors (WM) was involved in the production of an evidence gap map on PMM systems in primary care delivery in LMICs, we used their systematic search of various academic databases⁷ to inform this protocol’s preliminary PT ([Supplementary File 1](#) contains the MEDLINE search strategy). The combination of a scoping review of the PMM literature and the systematic search required by the evidence gap map mentioned above, helped us identify several social science theories that provide causal explanations about the mechanisms through which PMM interventions produce organizational change at multiple levels within a primary care system. A paper summarizing the findings from the evidence gap map will be published separately.

Context - We hypothesize that individual and interpersonal actions and reactions to SMI interventions will be influenced by the context in which providers, facilities and MOH organizational units are embedded, a feature that is particularly relevant in complex programs such as SMI^{46–48}. In this evaluation, context includes the institutional and policy setting that formalizes the laws and rules that govern the public sector in general, and the primary care delivery system, in particular. It also encompasses the internal organizational environment and the related practices, routines, and collective norms that drive organizational culture; and, the socio-economic local environments

Table 2. Summary of performance frameworks in El Salvador and Honduras.

Indicators	Baseline	Target	Indicators	Baseline	Target
EL SALVADOR			HONDURAS		
First Phase			First Phase		
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%
Second Phase			Second Phase		
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
Third Phase			Third Phase		
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

where primary care supply and demand interact for the production (or not) of outcomes (intended and otherwise). Finally, global agendas can also influence the choices and actions of high-level policy-makers^{49–51} possibly through their interactions within various policy networks^{52–56}.

Program actors' actions and reactions to context factors and to SMI interventions will likely vary according to their levels of interest, engagement, and resistance; relevant antecedents and experiences; the degree of system readiness among providers, managers and policy-makers; and on inherent features of program interventions and reforms, all of which have been empirically studied from the perspective of the theory of diffusion of innovations^{4,30,57–61}.

Mechanisms – We identified performance-driving mechanisms at three levels within a primary care system: individual, interpersonal and collective. From such multi-level perspective, primary care system states cannot solely be attributed to the behaviors of individuals but to the triggering of up to three types of interrelated causal mechanism: situational, action-oriented, and transformational mechanisms^{62–64}. Situational mechanisms refer to the macro, organizational-level environment in which system actors and their social interactions occur including, among others, the social institutions and collective norms that can exert influence on individual actors (macro-to-micro change). Action-oriented mechanisms explain how individual actors' ideas, actions and reactions influence other individuals' behaviors across the system, usually through diffusion from one to many actors (micro-micro change). Finally, transformational mechanisms explain how the sum of new behaviors by multiple actors bring about larger-scale changes in macro institutions and social norms (micro-to-macro transformation). In this evaluation, we propose to study individual and interpersonal mechanisms only.

At the individual level, we hypothesize that the motivation of healthcare providers has the potential to play a catalytic role in the generation of performance gains in facilities and local primary care delivery systems. At the interpersonal level, we theorize that social connections, imitation and the diffusion and dissemination of new beliefs and behaviors will further trigger their internalization and assimilation within primary care organizations. Downstream, the institutionalization of new organizational routines and practices through top-down policies and regulations will further normalize pro-performance behaviors across the primary care system. Program effects would accrue at any and all of these three levels of system change. Also, at each of these levels of potential system transformation, passive or active resistance by system actors may hinder, delay or entirely block the process of change, leading to any combinations of underperformance or performance failure.

Mainstream research in economics, psychology, organizational behavior, and public administration, among other fields, tends to assume that incentives and rewards serve as powerful motivators for the achievement of desirable behaviors among utility-maximizing, rational individuals^{65,66}. This approach has

fueled the design of various types of accountability-driven, PMM interventions that borrow performance approaches from the private sector and apply them in LMIC public settings, particularly in health and education. Such interventions attempt to reduce the misalignment of incentives between principals (voters, legislative bodies, executive-level leadership, funders, etc.) and their agents (program implementers, care providers, etc.),^{67–70}. Many public-sector reforms in LMICs and various global health partnerships have been influenced by this body of knowledge and by the adoption of PMM approaches by various global health program partnerships including SMI itself, the Global Finance Facility, GAVI, and the Global Fund to Fight AIDS, Tuberculosis and Malaria, among others.

Theoretical and empirical developments in public administration research also suggest that workforce motivation can be explained by intrinsic motives such as public service motivation, a socially learned set of preferences prevalent among individuals working in the public sector^{71,72}. In our search for substantive theories and evidence that could explain workers behaviors in primary care settings in LMICs, we decided to focus on self-determination theory, a macro theory of human motivation that has been used in recent years to study workforce motivation in various contexts, including LMIC^{44,73}. The theory has good cross-cultural validation and has demonstrated that individual workers who satisfy internal needs for competence, autonomy and relatedness feel intrinsically motivated and committed^{74–77}.

Regarding interpersonal mechanisms, we theorize that diffusion mechanisms within socially connected individuals can trigger interpersonal, action-oriented mechanisms that spread ideas, perceptions and behaviors from a few individuals to many more. We based this hypothesis on the diffusion of innovations theory^{59–61,78–81}, neo-institutional theory^{82–87} and, in particular, on recent characterizations of the processes of change triggered by PMM interventions in public sector organizations^{82,83}, and in healthcare settings^{88–90}.

Collective, whole-system change is the least theorized and empirically studied type of system transformation in healthcare. Given that we will study SMI at its mid-term, our assumption is that such types of transformational mechanisms may not be observable. However, based on the scarce number of studies that have addressed large-scale system change in healthcare settings^{4,30,91–93}, we propose that, were individual and interpersonal efforts at primary care system change sustain through time, such process may lead to the accumulation of new practices, routines and organizational behaviors beyond individual groups and teams through the norming of pro-performance, pro-social organizational cultures and social learning and modelling^{94–96}. Such system-level changes could lead to the emergence of quality and safety effects across the health primary care. The repetition of these cycles of improvement and learning would also lead to the generation of population-level health and equity outcomes, intended and otherwise.

Based on the evidence and theoretical framework discussed above, the preliminary PT was developed as a series of linked propositions (Box 1) and was also represented in graphical form

(Figure 2) to highlight the interrelated linkages between system elements and to avoid perceptions of linearity in causal reasoning.

Box 1. Preliminary PT Narrative

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 primary care 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 supply, 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: 1) 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); 2) 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); 3) 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, leading to the emergence of new organizational culture and collective norms (outcome); 4) 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) thus, increasing 5) the likelihood of triggering population-level health effects (outcome) and, potentially, 6) transforming the primary care system in a sustained fashion (outcome).

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. 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 and need to be considered in the characterization of context.

Data collection methods

Realist evaluation is method neutral. The nature of the phenomenon under study, the research questions, and the preliminary PT are the main factors that define study design and data collection methods^{31,32}. In this study protocol, the primary data collection methods will be key informant interviews, non-participatory

observation, and document review. Data collection will proceed between May 2017 and December 2018.

Study participants and sample. Study participants (or “key informants”) will be recruited based on their deep knowledge of and involvement in SMI, the central phenomenon of interest in the study^{45,97}. Therefore, while the sample size cannot be determined *a-priori*, for planning purposes we estimate the need to conduct approximately eighty (80) key informant interviews in the two countries. The adequacy of the final sample size will be continuously assessed during the research process.

Key informant interviews will be collected from four sets of actors: 1) Country policy- and program implementation actors; 2) Health care providers at primary care facilities; 3) Performance verification and evaluation stakeholders; and, 4) Program designers.

Country policy- and program implementation experts will have been involved in the governing of each country’s health system and/or in the design and implementation of phases 1 and 2 of the SMI program. These interviews will help understand the institutional and policy context; any relevant antecedents to the SMI program; and, also, concurrent investments by the government or external financing agencies in the same areas where SMI was implemented. Health care providers included will belong to high-performing primary care facilities directly involved in the delivery of health services in SMI areas of influence. Interviews will characterize the delivery of services; the relations between providers and the communities they serve; the features of the implementation of SMI; and, the perceptions and behaviors triggered by the latter. Performance verification, evaluation and initiative-wide management stakeholders (IADB and IHME) will be interviewed to acquire information about SMI’s interventions from their perspectives. Respondents will be invited to participate voluntarily in the study; and, no compensation will be provided.

Data collection. Specific questions will explore reasons for policy-makers to join SMI; respondent perceptions about the interventions under study; reactions triggered by the use of supply-side incentives and performance measurement and management arrangements and interventions; knowledge of and perceptions about context-specific factors that hinder or contribute to the actions and reactions among program actors, (local socio-economic context, institutional factors, and internal organizational environment); description of additional interventions that could explain program effects; and, effects or outcomes generated in an unintentional fashion. 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. Semi-structured interview guides will be used for data collection (Supplementary File 2).

We will use non-participant observation to collect information about the process of dissemination of results from the external measurement of performance at the end of phase 2, in

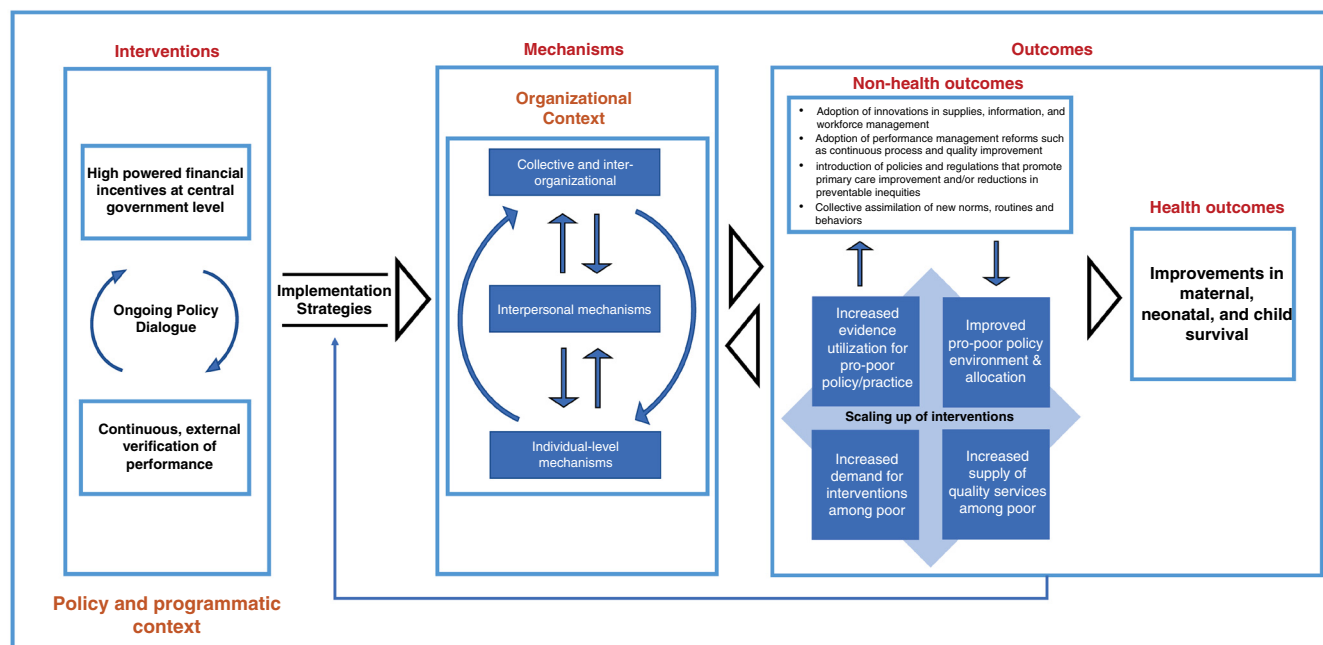


Figure 2. Preliminary program theory.

2018. We will document the process followed in the policy dialogue session, the agenda, components, objectives, and the reactions by domestic 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 also review key program documents pertinent to the design, implementation and evaluation of SMI in El Salvador and Honduras. Specific attention will be given to documents that describe the policy and program context in each country, the implementation strategies, and the performance and evaluation frameworks. Also, we will identify sources of secondary data that may be used for future triangulation. A complete list of reviewed documents will be maintained and included as a supplemental file with the final report of findings.

Data analysis

The analysis from the interviews will be conducted using an integrative methodology that merges both inductive and deductive approaches⁹⁸. We will construct a set of *a-priori* codes drawing from the theory-driven perspective used in realist evaluation to develop PTs, as described above. This will be combined with emergent inductive codes identified from a rigorous open coding process.

Realist evaluation collects data from various sources and, based on that, aims to build plausible accounts of key program events, adjustments in implementation, and on their intended and unintended effects⁹⁹. In an initial stage of data analysis, two coders will review a sub-set of transcripts in an iterative and systematic manner using the constant comparison method, and afterwards finalize the codebook through negotiation. Subsequent

transcripts will be coded by three experienced coders using the final codebook.

The coded data will be appraised using complementary analytic approaches. Coders will use iterative conceptual and pattern coding to identify major themes within and across cases. Within-case analysis will proceed as follows: deductive codes from each transcript will be aggregated into tables to identify preliminary linkages whereby certain outcomes are related to specific context-intervention-mechanism configurations. Coders will also scan each deductive coding category across the entire sample of responses to identify commonalities and differences, e.g. multiple combinations of contexts that could facilitate/inhibit the interventions; or a confluence of interventions that are catalytic and reinforce one another. We expect these 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 program interventions generated the observed effects, expected and otherwise.

At the conclusion of this stage, the resulting data will be integrated into preliminary analysis documents and diagrams to reflect the team's visualization of emergent, alternative theories of system change. The final thematic structure will be used to refine the preliminary program theory for developing within-case theories of system change at the delivery and policy levels. Also, for refining across-case theories at the same levels of analysis (e.g., policy- and primary care delivery-levels) to assess the extent to which the same mechanism may explain different outcomes in different contexts³⁰. Data analysis will be done with QRS nVivo. Furthermore, analysis will be complemented by contrasting evaluation results and emergent causal patterns

with other relevant SMI studies and data sets^{25–27,100,101}. The presentation of findings will be made following the standards developed for the reporting of realist evaluations¹⁰².

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. 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 affect the research process^{103–105}. 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.

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.

Discussion

This paper describes the protocol for a realist evaluation of PMM interventions introduced by SMI in the primary care systems in El Salvador and Honduras. The protocol proposes a contrasting case study design with outlier sampling, to understand how, why, and under what contextual conditions do SMI’s PMM interventions trigger high-levels of performance at scale. To our understanding, this is one of the few realist evaluations addressing PMM systems in primary care systems in LMICs. However, faced with budgetary and operational constraints, the researchers made design choices which generate challenges that are discussed below.

The first challenge relates the scarcity of theory-driven evidence on PMM systems initially identified when we scoped the health systems research literature in general, and in LMICs in particular. Many studies were undertheorized, disregarded context and mostly focused on dimensions of performance that addressed the effective delivery of specific health interventions. Various

impact evaluations tried to isolate the factors that contributed to the achievement of specific health outcomes, but under-theorized and infrequently measured the causal contributions of the PMM interventions of interest to this evaluation. Due to such design choices by conventional impact evaluations, many studies that have attempted to study primary care performance to date provide a poor understanding of the potential effects that PMM interventions may have on the behaviors of providers, facilities, or higher-levels of LMIC primary care systems. Similarly, if and when mechanisms or context factors are theorized or reported, such factors tended to be framed in terms of their contributions to health outcomes of interest, not to causal effects on the attitudes and behaviors of system actors at individual, interpersonal, or collective levels. Also, much of the health systems research tradition appears to have developed in isolation from other disciplines that have a rich tradition of PMM theorizing, research and practice including public administration, organizational studies, and social psychology, among others. To address the issues above, we decided to, first, develop a conceptual framework that is informed by multi-disciplinary research and theory mostly arising from experiences in industrialized settings.

Another challenge arose from the choice of a cross-sectional study design. While it would have been desirable to accompany SMI and its domestic partners in a continuous process of sense-making and performance information use, such approach was not feasible with the resources available. Therefore, this evaluation may serve as the starting point in theorizing multi-level, complex and dynamic processes of large-scale system change in high-performing primary care systems in SMI; we also expect that its findings will further support and complement SMI’s learning and evaluation plans, and contribute to generating theory and evidence of relevance to other global contexts.

Issues of country case selection also need to be explicitly addressed. As this is the first realist evaluation done in the SMI context, the research team chose to focus on characterizing high-performance systems and thus only included positive outliers at the national and primary care delivery levels. Such decision was made to maximize information power and richness from studying similar, extreme cases. While acknowledging that a contrasting case of high- and low-performer countries and/or primary care delivery systems would have been desirable, such design was not feasible at this stage, yet can be undertaken in the future. Furthermore, understanding how and why system improvements sustain (or not) through time remains a valuable yet complex research endeavor that requires further theorizing and additional empirical studies that are outside the scope of this evaluation. Given the unusual duration of SMI’s implementation period, the initiative offers a unique learning space from which to acquire new knowledge about the processes underlying system inertia, resistance to change, and the dynamics of systems that ‘learn’ and adapt through time^{106–114}.

The refined PT and other results from this evaluation have several anticipated uses and applications. For instance, we expect that program implementers will use the findings to assess program adjustments in its third and final phase (2018–2020);

also, to identify options for re-designing domestic health policies and new evaluation priorities; and, to inform the design of longitudinal, experimental or quasi-experimental evaluation designs that may deepen one or more of the various casual patterns identified.

There is growing concern that high-performing primary care systems are needed to prevent global pandemics and to deliver on the promises of Universal Health Coverage and the Sustainable Development Goals². This realist evaluation aims to contribute to such ambitious goals by conducting this study in El Salvador and Honduras, two of the top-performing low-income countries in SMI. The findings of this evaluation in regard to how, why and under what conditions have these two low-income countries transformed their primary care PMM systems, will

provide learning opportunities for spreading insights, evidence and new theory to other countries trying to address similar challenges.

Data availability

No data is associated with this article.

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: MEDLINE search strategy

[Click here to access the data](#)

Supplementary File 2: In-depth interview guides.

[Click here to access the data](#)

References

- 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](#)
- Gates B: **The next epidemic—lessons from Ebola.** *N Engl J Med.* 2015; **372**(15): 1381–4. [PubMed Abstract](#) | [Publisher Full Text](#)
- Kruk ME, Gage AD, Arsenault C, *et al.*: **High-quality health systems in the Sustainable Development Goals era: time for a revolution.** *Lancet Glob Health.* 2018; pii: S2214-109X(18)30386-3. [PubMed Abstract](#) | [Publisher Full Text](#)
- 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](#)
- Borgonovi E, Anessi-Pessina E, Bianchi C: **Outcome-Based Performance Management in the Public Sector.** Cham, Switzerland: Springer; 2017. [Publisher Full Text](#)
- Rajala T, Laiho H, Vakkuri J: **Shifting from Output to Outcome Measurement in Public Administration—Arguments Revisited.** In: Borgonovi E, Anessi-Pessina E, Bianchi C, eds. *Outcome-Based Performance Management in the Public Sector.* Cham, UK: Springer; 2018; 3–23. [Publisher Full Text](#)
- Munar W, Snilstveit B, Stevenson J, *et al.*: **Evidence gap map of performance measurement and management in primary care delivery systems in low- and middle-income countries - Study protocol [version 1; referees: 2 approved].** *Gates Open Res.* 2018; **2**: 27. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bevan G: **Setting targets for health care performance: lessons from a case study of the English NHS.** *Natl Inst Econ Rev.* 2006; **197**(1): 67–79. [Publisher Full Text](#)
- Bevan G, Hood C: **What's measured is what matters: targets and gaming in the English public health care system.** *Public Adm.* 2006; **84**(3): 517–38. [Publisher Full Text](#)
- Bevan G, Wilson D: **Does 'naming and shaming' work for schools and hospitals? Lessons from natural experiments following devolution in England and Wales.** *Public Money Manage.* 2013; **33**(4): 245–52. [Publisher Full Text](#)
- Suthar AB, Nagata JM, Nsanzimana S, *et al.*: **Performance-based financing for improving HIV/AIDS service delivery: a systematic review.** *BMC Health Serv Res.* 2017; **17**(1): 6. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Pollitt C: **Performance management 40 years on: a review. Some key decisions and consequences.** *Public Money Manage.* 2018; **38**(3): 167–74. [Publisher Full Text](#)
- Cepiku D, Hinna A, Scarozza D, *et al.*: **Performance information use in public administration: an exploratory study of determinants and effects.** *Journal of Management & Governance.* 2017; **21**(4): 963–91. [Publisher Full Text](#)
- Belle N, Cantarelli P: **What Causes Unethical Behavior? A Meta-Analysis to Set an Agenda for Public Administration Research.** *Public Adm Rev.* 2017; **77**(3): 327–39. [Publisher Full Text](#)
- Kelman S, Friedman JN: **Performance improvement and performance dysfunction: an empirical examination of distortionary impacts of the emergency room wait-time target in the English National Health Service.** *J Public Adm Res Theory.* 2009; **19**(4): 917–46. [Publisher Full Text](#)
- 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](#)
- Pollitt C: **The logics of performance management.** *Evaluation.* 2013; **19**(4): 346–63. [Publisher Full Text](#)
- Effective Practice and Organisation of Care (EPOC): **The EPOC taxonomy of health systems interventions.** EPOC Resources for review author. Oslo, Norway:

Norwegian Knowledge Centre for the Health Services, 2016.

Reference Source

19. Pantoja T, Opiyo N, Ciapponi A, *et al.*: **Implementation strategies for health systems in low-income countries: an overview of systematic reviews (Protocol).** *Cochrane Database Syst Rev*. The Cochrane Library. 2014; (5). [PubMed Abstract](#) | [Publisher Full Text](#)
20. Wiyongse CS, Paulsen E, Lewin S, *et al.*: **Financial arrangements for health systems in low-income countries: an overview of systematic reviews.** *Cochrane Database Syst Rev*. 2017; (9): CD011084. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
21. Ivers N, Jamtvedt G, Flottorp S, *et al.*: **Audit and feedback: effects on professional practice and healthcare outcomes.** *Cochrane Database Syst Rev*. 2012; (6): CD000259. [PubMed Abstract](#) | [Publisher Full Text](#)
22. Ivers NM, Grimshaw JM, Jamtvedt G, *et al.*: **Growing literature, stagnant science? Systematic review, meta-regression and cumulative analysis of audit and feedback interventions in health care.** *J Gen Intern Med*. 2014; 29(11): 1534–41. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Molina E, Carella L, Pacheco A, *et al.*: **Community monitoring interventions to curb corruption and increase access and quality in service delivery: a systematic review.** *J Dev Effect*. 2017; 9(4): 462–99. [Publisher Full Text](#)
24. Eichler R, Nelson J, Iriarte E, *et al.*: **The initial prize in the Salud Mesoamerica initiative results-based aid initiative - Strengthened Health Systems for Reproductive, Maternal, Neonatal and Child Outcomes.** Washington, DC: Inter-American Development Bank, 2017. [Publisher Full Text](#)
25. Mokdad AH, Gagnier MC, Colson KE, *et al.*: **Missed Opportunities for Measles, Mumps, and Rubella (MMR) Immunization in Mesoamerica: Potential Impact on Coverage and Days at Risk.** *PLoS One*. 2015; 10(10): e0139680. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
26. Mokdad AH, Colson KE, Zúñiga-Brenes P, *et al.*: **Salud Mesoamérica 2015 Initiative: design, implementation, and baseline findings.** *Popul Health Metr*. 2015; 13(1): 3. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
27. Mokdad AH, Gagnier MC, Colson KE, *et al.*: **Health and wealth in Mesoamerica: findings from Salud Mesoamérica 2015.** *BMC Med*. 2015; 13(1): 164. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
28. 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](#)
29. 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](#)
30. 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](#)
31. Pawson R, Tilley N: **Realistic evaluation.** Sage; 1997. [Reference Source](#)
32. Pawson R: **Evidence-based policy: A realist perspective.** Thousand Oaks, CA: Sage Publications; 2006. [Reference Source](#)
33. Adams A, Sedalia S, McNab S, *et al.*: **Lessons learned in using realist evaluation to assess maternal and newborn health programming in rural Bangladesh.** *Health Policy Plan*. 2016; 31(2): 267–75. [PubMed Abstract](#) | [Publisher Full Text](#)
34. Blaise P, Kegels G: **A realistic approach to the evaluation of the quality management movement in health care systems: a comparison between European and African contexts based on Mintzberg's organizational models.** *Int J Health Plann Manage*. 2004; 19(4): 337–64. [PubMed Abstract](#) | [Publisher Full Text](#)
35. Gilmore B, McAuliffe E, Larkan F, *et al.*: **How do community health committees contribute to capacity building for maternal and child health? A realist evaluation protocol.** *BMJ Open*. 2016; 6(11): e011885. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
36. Hernández AR, Hurtig AK, Dahlblom K, *et al.*: **More than a checklist: a realist evaluation of supervision of mid-level health workers in rural Guatemala.** *BMC Health Serv Res*. 2014; 14: 112. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
37. Kwame A, van Dijk H, Agyepong IA: **Advancing the application of systems thinking in health: realist evaluation of the Leadership Development Programme for district manager decision-making in Ghana.** *Health Res Policy Syst*. 2014; 12(1): 29. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
38. Maluka S, Kamuzora P, SanSebastián M, *et al.*: **Implementing accountability for reasonableness framework at district level in Tanzania: a realist evaluation.** *Implement Sci*. 2011; 6(1): 11. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
39. Marchal B, Dedzo M, Kegels G: **A realist evaluation of the management of a well-performing regional hospital in Ghana.** *BMC Health Serv Res*. 2010; 10(1): 24. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
40. Mirzoev T, Etiaba E, Ebenso B, *et al.*: **Study protocol: realist evaluation of effectiveness and sustainability of a community health workers programme in improving maternal and child health in Nigeria.** *Implement Sci*. 2016; 11(1): 83. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
41. Prashanth NS, Marchal B, Devadasan N, *et al.*: **Advancing the application of systems thinking in health: a realist evaluation of a capacity building programme for district managers in Tumkur, India.** *Health Res Policy Syst*. 2014; 12(1): 42. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
42. Prashanth NS, Marchal B, Kegels G, *et al.*: **Evaluation of capacity-building program of district health managers in India: a contextualized theoretical framework.** *Front Public Health*. 2014; 2: 89. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
43. van de Klundert J, van Dongen-van den Broek J, Yesuf EM, *et al.*: **'We are planning to leave, all of us'-a realist study of mechanisms explaining healthcare employee turnover in rural Ethiopia.** *Hum Resour Health*. 2018; 16(1): 37. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
44. Vareilles G, Marchal B, 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.** *BMJ Open*. 2015; 5(11): e008614. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
45. Patton MQ: **Qualitative Research & Evaluation Methods: Integrating Theory and Practice.** 4th ed. Thousand Oaks, CA: Sage Publications; 2014. [Reference Source](#)
46. Bourne M, Franco-Santos M, Micheli P, *et al.*: **Performance measurement and management: a system of systems perspective.** *Int J Prod Res*. 2018; 56(8): 2788–99. [Publisher Full Text](#)
47. Kok MC, Broerse JEW, Theobald S, *et al.*: **Performance of community health workers: situating their intermediary position within complex adaptive health systems.** *Hum Resour Health*. 2017; 15(1): 59. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
48. Kok MC, Kane SS, Tulloch O, *et al.*: **How does context influence performance of community health workers in low- and middle-income countries? Evidence from the literature.** *Health Res Policy Syst*. 2015; 13: 13. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
49. Shiffman J: **Generating political priority for maternal mortality reduction in 5 developing countries.** *Am J Public Health*. 2007; 97(5): 796–803. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
50. Shiffman J, Schmitz HP, Berlan D, *et al.*: **The emergence and effectiveness of global health networks: findings and future research.** *Health Policy Plan*. 2016; 31 Suppl 1: i110–23. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
51. Hafner T, Shiffman J: **The emergence of global attention to health systems strengthening.** *Health Policy Plan*. 2013; 28(1): 41–50. [PubMed Abstract](#) | [Publisher Full Text](#)
52. Hulton L, Matthews Z, Martin-Hilber A, *et al.*: **Using evidence to drive action: a "revolution in accountability" to implement quality care for better maternal and newborn health in Africa.** *Int J Gynaecol Obstet*. 2014; 127(1): 96–101. [PubMed Abstract](#) | [Publisher Full Text](#)
53. Weyland K: **Bounded rationality and policy diffusion: social sector reform in Latin America.** Princeton University Press; 2009. [Reference Source](#)
54. Weyland K: **Theories of Policy Diffusion Lessons from Latin American Pension Reform.** *World Polit*. 2005; 57(2): 269–95. [Publisher Full Text](#)
55. Smith SL, Shiffman J: **Setting the global health agenda: The influence of advocates and ideas on political priority for maternal and newborn survival.** *Soc Sci Med*. 2016; 166: 86–93. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
56. Shiffman J: **Network advocacy and the emergence of global attention to newborn survival.** *Health Policy Plan*. 2016; 31 Suppl 1: i60–73. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
57. Greenhalgh T, Robert G, Macfarlane F, *et al.*: **Diffusion of innovations in service organizations: systematic review and recommendations.** *Milbank Q*. 2004; 82(4): 581–629. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
58. 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](#)
59. 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](#)
60. McMullen H, Griffiths C, Leber W, *et al.*: **Explaining high and low performers in complex intervention trials: a new model based on diffusion of innovations**

theory. *Trials*. 2015; **16**: 242.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

61. Rogers EM: **Diffusion of Innovations**. Fifth ed. New York: Free Press; 2003.
[Reference Source](#)
62. Hedström P, Ylikoski P: **Causal mechanisms in the social sciences**. *Annu Rev Sociol*. 2010; **36**: 49–67.
[Publisher Full Text](#)
63. 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](#)
64. Hedström P, Wennberg K: **Causal mechanisms in organization and innovation studies**. *Innovation*. 2017; **19**(1): 91–102.
[Publisher Full Text](#)
65. Elster J, editor: **Rational choice**. New York: NYU Press; 1986.
[Reference Source](#)
66. Monroe KR, Maher KH: **Psychology and rational actor theory**. *Polit Psychol*. 1995; **16**(1): 1–21.
[Publisher Full Text](#)
67. 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](#)
68. Grossman SJ, Hart OD: **An analysis of the principal-agent problem**. *Econometrica*. 1983; **51**(1): 7–45.
[Publisher Full Text](#)
69. 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](#)
70. Eisenhardt KM: **Agency theory: An assessment and review**. *Acad Manage Rev*. 1989; **14**(1): 57–74.
[Publisher Full Text](#)
71. Perry JL, Wise LR: **The motivational bases of public service**. *Public Adm Rev*. 1990; **50**(3): 367–73.
[Publisher Full Text](#)
72. Perry JL, Hondeghem A, Wise LR: **Revisiting the motivational bases of public service: Twenty years of research and an agenda for the future**. *Public Adm Rev*. 2010; **70**(5): 681–90.
[Publisher Full Text](#)
73. Vareilles G, Pommier J, Marchal B, et al.: **Understanding the performance of community health volunteers involved in the delivery of health programmes in underserved areas: a realist synthesis**. *Implement Sci*. 2017; **12**(1): 22.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
74. Deci EL, Ryan RM: **Intrinsic motivation and self-determination in human behavior**. New York: Plenum; 1985.
[Publisher Full Text](#)
75. Gagné M, Deci EL: **Self-determination theory and work motivation**. *J Organ Behav*. 2005; **26**(4): 331–62.
[Publisher Full Text](#)
76. Deci EL, Ryan RM: **Self-determination theory: A macrotheory of human motivation, development, and health**. *Can Psychol Psychol Canadienne*. 2008; **49**(3): 182–5.
[Publisher Full Text](#)
77. Ryan RM, Deci EL: **Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being**. *Am Psychol*. 2000; **55**(1): 68–78.
[PubMed Abstract](#) | [Publisher Full Text](#)
78. Greenhalgh T, Wherton J, Papoutsis C, et al.: **Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies**. *J Med Internet Res*. 2017; **19**(11): e367.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
79. Green LW, Ottoson JM, García C, et al.: **Diffusion theory and knowledge dissemination, utilization, and integration in public health**. *Annu Rev Public Health*. 1993; **30**: 151–74.
[PubMed Abstract](#) | [Publisher Full Text](#)
80. Ottoson JM: **Knowledge-for-action theories in evaluation: knowledge utilization, diffusion, implementation, transfer and translation**. *New Dir Eval*. 2009; **2009**(124): 7–20.
[Publisher Full Text](#)
81. Ottoson JM, Hawe P: **New Directions for Evaluation - Knowledge Utilization, Diffusion, Implementation, Transfer, and Translation: Implications for Evaluation**. *New Dir Eval*. 2009; **2009**(124): 3–100.
[Reference Source](#)
82. Ospina S, Cunill Grau N, Zaltsman A: **Performance evaluation, public management improvement and democratic accountability**. *Public Manag Rev*. 2004; **6**(2): 229–51.
[Publisher Full Text](#)
83. Cunill-Grau N, Ospina SM: **Performance measurement and evaluation systems: Institutionalizing accountability for governmental results in Latin America**. *New Dir Eval*. 2012; **2012**(134): 77–91.
[Publisher Full Text](#)
84. Ospina S, Cunill Grau N: **Institutionalizing Accountability for Governmental Results: Public Performance Measurement and Evaluation Systems in Latin America**. *Public Management Research Association Conference*; 2011; 57.
[Reference Source](#)
85. Scott WR, Ruef M, Mendel PJ, et al.: **Institutional change and healthcare organizations: From professional dominance to managed care**. Chicago: The University of Chicago Press; 2000.
[Reference Source](#)
86. Rautiainen A, Järvenpää M: **Institutional logics and responses to performance measurement systems**. *Financial Accountability & Management*. 2012; **28**(2): 164–88.
[Publisher Full Text](#)
87. Thornton PH, Ocasio W, Lounsbury M: **The institutional logics perspective: A new approach to culture, structure, and process**. Oxford University Press; 2012.
[Reference Source](#)
88. Atun RA, Kyrtatis I, Jelic G, et al.: **Diffusion of complex health innovations--implementation of primary health care reforms in Bosnia and Herzegovina**. *Health Policy Plan*. 2007; **22**(1): 28–39.
[PubMed Abstract](#) | [Publisher Full Text](#)
89. Bradley EH, Curry LA, Taylor LA, et al.: **A model for scale up of family health innovations in low-income and middle-income settings: a mixed methods study**. *BMJ Open*. 2012; **2**(4): pii: e000987
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
90. Bradley EH, Byam P, Alpern R, et al.: **A systems approach to improving rural care in Ethiopia**. *PLoS One*. 2012; **7**(4): e35042.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
91. 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](#)
92. Greenhalgh T, Macfarlane F, Barton-Sweeney C, et al.: **"If we build it, will it stay?" A case study of the sustainability of whole-system change in London**. *Milbank Q*. 2012; **90**(3): 516–47.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
93. Perla RJ, Bradbury E, Gunther-Murphy C: **Large-scale improvement initiatives in healthcare: a scan of the literature**. *J Healthc Qual*. 2013; **35**(1): 30–40.
[PubMed Abstract](#) | [Publisher Full Text](#)
94. Kincaid DL: **From innovation to social norm: bounded normative influence**. *J Health Commun*. 2004; **9** Suppl: 37–57.
[PubMed Abstract](#) | [Publisher Full Text](#)
95. Buchanan D, Fitzgerald L, Ketley D, et al.: **No Going Back: A Review of the Literature on Sustaining Organizational Change**. *Int J Manag Rev*. 2005; **7**(3): 189–205.
[Publisher Full Text](#)
96. Lanham HJ, McDaniel RR Jr, Crabtree BF, et al.: **How improving practice relationships among clinicians and nonclinicians can improve quality in primary care**. *Jt Comm J Qual Patient Saf*. 2009; **35**(9): 457–66.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
97. Malterud K, Siersma VD, Guassora AD: **Sample Size in Qualitative Interview Studies: Guided by Information Power**. *Qual Health Res*. 2016; **26**(13): 1753–60.
[PubMed Abstract](#) | [Publisher Full Text](#)
98. 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](#)
99. George AL, Bennett A: **Case Studies and Theory Development in the Social Science**. Cambridge, MA: MIT Press; 2005.
[Reference Source](#)
100. Colombara DV, Hernández B, Gagnier MC, et al.: **Breastfeeding Practices among Poor Women in Mesoamerica**. *J Nutr*. 2015; **145**(8): 1958–65.
[PubMed Abstract](#) | [Publisher Full Text](#)
101. El Bcheraoui C, Palmisano EB, Dansereau E, et al.: **Healthy competition drives success in results-based aid: Lessons from the Salud Mesoamérica Initiative**. *PLoS One*. 2017; **12**(10): e0187107.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
102. 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](#)
103. Reynolds J, DiLiberto D, Mangham-Jefferies L, et al.: **The practice of 'doing' evaluation: lessons learned from nine complex intervention trials in action**. *Implement Sci*. 2014; **9**: 75.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
104. Barry CA, Britten N, Barber N, et al.: **Using reflexivity to optimize teamwork in qualitative research**. *Qual Health Res*. 1999; **9**(1): 26–44.
[PubMed Abstract](#) | [Publisher Full Text](#)
105. Finlay L: **Negotiating the swamp: the opportunity and challenge of reflexivity in research practice**. *Qual Res*. 2002; **2**(2): 209–30.
[Publisher Full Text](#)
106. McGinnis JM, Stuckhardt L, Saunders R, et al.: **Best Care at Lower Cost: The Path to Continuously Learning Health Care in America**. Institute of Medicine of the National Academies: National Academies Press; 2013.
[PubMed Abstract](#) | [Publisher Full Text](#)
107. Laihoen H: **A managerial view of the knowledge flows of a health-care system**.

Knowl Man Res Pract. 2015; **13**(4): 475–85.

[Publisher Full Text](#)

108. Argyris C, Schon DA: **Organizational Learning: A theory of action approach.** Reading, MA: Addison-Wesley; 1978.
[Reference Source](#)
109. Crossan MM, Lane HW, White RE: **An Organizational Learning Framework: From Intuition to Institution.** *Acad Manage Rev.* 1999; **24**(3): 522–37.
[Publisher Full Text](#)
110. Moynihan DP: **Goal-based learning and the future of performance management.** *Public Adm Rev.* 2005; **65**(2): 203–16.
[Publisher Full Text](#)
111. 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](#)

112. Teece DJ, Pisano G, Shuen A: **Dynamic capabilities and strategic management.** *Strateg Manage J.* 1997; **18**(7): 509–33.
[Publisher Full Text](#)
113. Rothaermel FT, Hess AM: **Building dynamic capabilities: Innovation driven by individual-, firm-, and network-level effects.** *Organ Sci.* 2007; **18**(6): 898–921.
[Publisher Full Text](#)
114. Hovmand PS, Gillespie DF: **Implementation of evidence-based practice and organizational performance.** *J Behav Health Serv Res.* 2010; **37**(1): 79–94.
[PubMed Abstract](#) | [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

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

² 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 3rd paragraph under *Study Setting*: the 1st 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.
