STUDY PROTOCOL
Evidence gap map of performance measurement and management in primary care delivery systems in low- and middle-income countries – Study protocol [version 1; peer review: 2 approved]

Wolfgang Munar, Birte Snilstveit, Jennifer Stevenson, Nilakshi Biswas, John Eyers, Gisela Butera, Theresa Baffour, Ligia E. Aranda

1Milken Institute School of Public Health, Department of Global Health, George Washington University, Washington, DC, 20052, USA
2International Initiative for Impact Evaluation (3Ie), London International Development Centre, London, WC1H 0PD, UK

Abstract
Background. For the last two decades there has been growing interest in governmental and global health stakeholders about the role that performance measurement and management systems can play for the production of high-quality and safely delivered primary care services. Despite recognition and interest, the gaps in evidence in this field of research and practice in low- and middle-income countries remain poorly characterized. This study will develop an evidence gap map in the area of performance management in primary care delivery systems in low- and middle-income countries.

Methods. The evidence gap map will follow the methodology developed by 3Ie, the International Initiative for Impact Evaluation, to systematically map evidence and research gaps. The process starts with the development of the scope by creating an evidence-informed framework that helps identify the interventions and outcomes of relevance as well as help define inclusion and exclusion criteria. A search strategy is then developed to guide the systematic search of the literature, covering the following databases: Medline (Ovid), Embase (Ovid), CAB Global Health (Ovid), Cinahl (Ebsco), Cochrane Library, Scopus (Elsevier), and Econlit (Ovid). Sources of grey literature are also searched. Studies that meet the inclusion criteria are systematically coded, extracting data on intervention, outcome, measures, context, geography, equity, and study design. Systematic reviews are also critically appraised using an existing standard checklist. Impact evaluations are not appraised but will be coded according to study design. The process of map-building ends with the creation of an evidence gap map graphic that displays the available evidence according to the intervention and outcome framework of interest.

Discussion. Applications arising from the evidence map will be discussed in a separate paper that will summarize findings and make recommendations for the development of a prioritized research agenda.
Keywords
Accountability, Evidence gap maps, Implementation strategies, Low- and middle-income countries, Performance measurement and management systems, Primary Care delivery systems, Quality of care, Safety

Corresponding author: Wolfgang Munar (wolfgangmunar@email.gwu.edu)

Author roles: Munar W: Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; Snilstveit B: Conceptualization, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; Stevenson J: Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; Biswas N: Investigation, Writing – Review & Editing; Eyers J: Methodology, Writing – Review & Editing; Butera G: Methodology; Baffour T: Investigation, Project Administration, Writing – Review & Editing; Aranda LE: Investigation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: Bill and Melinda Gates Foundation [OPP1154415, OPP1149078]. This work was supported by the Bill & Melinda Gates Foundation through grants to George Washington University [OPP1154415], and Ariadne Labs through Brigham and Women’s Hospital [OPP1149078].

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 Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


First published: 29 May 2018, 2:27 (https://doi.org/10.12688/gatesopenres.12826.1)
Background
The critical role that primary care delivery systems can play in helping achieve desirable societal goals in low- and middle-income countries (LMIC) has been widely recognized. Given their potential to serve as first points of contact for continuous, coordinated, comprehensive and people-centered health services, high-performing primary care systems are a necessary element for the achievement of the sustainable development goals, the operationalization of calls for universal health coverage, and the management of global pandemics. While considerable research is available on primary health care and its constitutive elements, it is not clear which approaches are most effective to ensure that primary care systems consistently deliver safe and quality services, that harmful services are not delivered, that primary care delivery systems acquire the capabilities required for continuous improvement, and that all of the above add up to improved population health and other socially valued outcomes.

The objective of this study is to identify and describe existing evidence on the effects of interventions in the area of performance measurement and management in primary care delivery systems in LMICs and, also, to provide easy access to such evidence for relevant decision makers. The resulting evidence gap map (EGM) will inform the development of a prioritized research agenda for primary care delivery systems in LMIC.

Why is this study relevant to research, policy and practice in LMICs?
There are multiple approaches, frameworks, and conceptualizations for characterizing health systems, measuring and managing their performance, and typifying health system interventions. The study uses a multidisciplinary approach to identify and characterize the relevant literature from different fields and disciplines such as organizational science, health systems research, and public-sector management.

Organizational performance refers to the results generated by an organization and measured against its intended aims. In private sector organizations, performance can be a function of profits, organizational efficiency and effectiveness, quality of goods and services, market share, and customer satisfaction. In public administration, the definition of organizational performance has evolved with the changing framings for the role of the State in the production and delivery of public goods and services. Historically, governments initially emphasized aspects of performance such as the control of inputs and the compliance with standards. Subsequent framings shifted, first, towards a focus on the quantity and quality of outputs, productivity, and efficiency and, in recent years, to outcomes and policy impacts and, in the case of the health sector, to social values like patient- or people-centered health services.

On the research side, the theories of organizational performance have followed, in general, the evolution of the practice of performance management in high-income countries. According to Talbot, an initial set of theories and frameworks were focused on characterizing associations between individual elements of performance and organizational effectiveness. Afterwards, researchers focused on excellence, quality and organizational culture which led to the development of a first wave of models of performance measurement and management. These models did not account for differences between public and private sector dimensions of performance but were nonetheless adopted by governments around the world. In the 90s, the focus shifted from theoretical perspectives about organizational performance to interest in how to measure goal achievement in public and private sector organizations using performance models such as the Balance Scorecard and others. Interest in performance measurement and management spread around the world, and international comparisons and benchmarking of performance flourished in various sectors such as governance, health, and education.

During the last 40 years, innovations in performance measurement and management in the health sector have been prevalent in the United Kingdom, Canada, Australia, New Zealand, Sweden and the US, among other countries. Amplified by multilateral finance organizations and some bilateral agencies, performance measurement and management systems have spread among LMIC, sometimes as central aspects of large-scale public-sector reforms and, also, as stand-alone health sector reforms. Some elements of performance measurement and management have spread more than others particularly performance-based financing, pay-for-performance, performance budgeting and contracting and the use of financial incentives (defined below).

The spread of the practice and research of performance measurement and management has also affected the global health architecture and its governance. The interest among donor governments, multilateral finance institutions, bilateral agencies, and global philanthropies started shifting since the late 90s from a focus on funding inputs towards an interest on the production of measurable results aid effectiveness, and accountability. Such shifts in preferences contributed to the emergence of new global organizational forms and partnerships such as the Global Fund to Fight HIV/AIDS, Tuberculosis and Malaria, GAVI the Vaccine Alliance, the Global Finance Facility, and the Mesoamerican Health Facility, to name a few.

In terms of effectiveness, 40 years of research on performance measurement and management have shown that, despite many challenges, such systems can be effective. There is evidence, also, of the generation of unintended effects in the public and private sectors.

In the area of health systems research, research conducted to date by the Cochrane Collaboration has generated approximately 200 systematic reviews addressing the effective organization of health services. While the majority of these have been focused on issues of relevance to high-income countries research and policy, there is a growing portfolio of reviews focused on delivery and financial arrangements, as well as implementation strategies in LMIC.

The components of a generic performance management system are delineated in Figure 1 and represent an adaptation of two
frameworks. The first, is a generic framework of public management in public sector organizations developed by Pollitt and the second is the result of ongoing research on primary care performance in Mesoamerica led by one of this study’s authors (WM). Framework components include: 1) A context in which various policies, organizations, programs and health interventions coexist with system actors and stakeholders; 2) one or more performance management interventions; 3) activities for measuring the results from the implementation of primary care policies and programs, and its ensuing data; 4) a process through which raw performance data is made sense of and transformed into performance information; 5) dissemination of performance information among system actors and stakeholders with the intent of making it actionable; 6) performance information use, misuse or non-use; 7) implementation of planned action based on the use of performance information; and, 8) the effects from the implementation of planned action and clinical and managerial improvements (proximal processes, outputs and outcomes, and distal, societal and population-level outcomes).

The production of actual, measurable performance is a complex, dynamic phenomenon. Real performance can be very hard if at all possible to observe. Its measurement is characterized by lags between the introduction of interventions, the production of effects, and their measurement. These delays can also create a disconnect between action, measurement and results. Once measurement occurs and performance information is available, system actors and stakeholders can respond to the perceived performance gap by using, not using, or misusing such information. To be effective, performance information needs to apprise subsequent organizational action. Changes in strategic direction or operational tactics would also have to be effectively implemented for outcomes to be generated.

Based on the above theoretical and practice-oriented considerations, the study defines performance measurement and management in a primary care delivery system as the introduction of management systems focused on measuring organizational processes, outputs and outcomes with the proximal aim of informing the introduction of clinical, managerial, programmatic and policy changes and the ultimate goal of contributing to socially valued, population level health outcomes.

**Intervention and outcomes of interest**

To typify interventions and outcomes, this study uses an adapted version of the taxonomy developed by the Effective Practice and Organization of Care (EPOC). Within the general categories described in such taxonomy, the study will focus on two: 1) Implementation strategies and, in a limited way, 2) Financial arrangements. Implementation strategies are defined as interventions designed to bring about changes in healthcare organization, the behavior of healthcare professionals or the use of health services by recipients. Financial arrangements refer to changes in how funds are collected, insurance schemes, how services are purchased, and the use of targeted financial incentives or disincentives. These two categories of interventions can operate at individual- (providers, managers, etc.) or organizationallevels (facilities, networks of care, local health systems, etc.).

Accountability interventions at individual, organizational, and community-levels were also included as a separate category. Given the growing interest on values like patient-centered care
and the confluence of the latter with long-standing interest in community participation and citizen engagement, there has been an increase in the availability of evidence surrounding the policy relevance of social accountability interventions as a system of external control that can drive performance improvements in primary care delivery systems. There has also been a long-standing focus of government-driven performance reforms focused on inducing accountability among healthcare providers using internal accountability interventions such as audit and feedback, supervision, and others. For the purposes of this study, accountability arrangements are defined as the organizational and institutional arrangements used by system actors within governments to steward the delivery of public services towards increased performance.

In summary, a total of 11 interventions in three major categories were selected for inclusion in the categories of implementation strategies, accountability arrangements, and financial arrangements. Table 1 summarizes the interventions included in the study.

Implementation strategies
In this category, the study identified 8 interventions of relevance including: 1) In-service training, a form of positive behavior support aimed at increasing the capabilities of individual primary care system actors; 2) Reminders, manual or computerized interventions that prompt individual providers to perform an action during a clinical exchange and can include, among others, job aids, paper reminders, checklists, and computer decision support systems; 3) Clinical practice guidelines, or systematically developed statements to assist healthcare providers and patients to decide on appropriate health care for specific circumstances; 4) Continuous education, referring to courses, workshops, or other educational meetings aimed at increasing the technical competencies of primary care providers; 5) Clinical incident reporting, or systems for reporting critical incidents and adverse or undesirable effects as a means to improving the safety of healthcare delivery; 6) Local opinion leaders, referring to the identification and use of identifiable local opinion leaders to promote good clinical practices; 7) Continuous quality improvement defined as the iterative process to review and improve care that includes involvement of healthcare teams, analysis of a process or system, a structured process improvement method or problem-solving approach, and use of data analysis to assess changes. It will include lean management as one of the approaches used to improve efficiency and quality in service provider organizations; and, 8) Supervision, defined as routine control visits by senior primary care staff to providers and facilities.

Accountability arrangements
The interventions included in this EGM are: 1) Audit and feedback, defined as a summary of primary care provider or facility performance over a specified period of time, given in a written, electronic, or verbal format; such interventions can occur at individual provider as well as at organizational, facility level; 2) Public release of performance data, defined as arrangements to inform the public about the performance of primary care providers or facilities in written or electronic formats; and, 3) Social accountability interventions, defined as an accountability arrangement in which community members and/or civil society organizations are involved in the monitoring of performance of primary care providers or facilities.

Financial arrangements
There are many variations in this type intervention and contested definition among them. The interventions of interest to this study are under the general heading of Performance-Based Financing (PBF) but can also include Results-Based Financing (RBF), Pay-for-Performance (PFP), and the use of provider rewards and incentives. For precision purposes, we include the definitions developed by Musgrove for these terms:

- Results-based financing refers to any program that rewards the delivery of one or more outputs or outcomes by one or more incentives, financial or otherwise, upon verification that the agreed-upon result has actually been delivered. Incentives may be directed to service providers (supply side), program beneficiaries (demand side) or both. Payments or other rewards are not used for recurrent inputs.

<table>
<thead>
<tr>
<th>Intervention categories</th>
<th>Individual-level provider interventions</th>
<th>Organizational-level interventions</th>
<th>Societal, community-level interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implementation strategies</td>
<td>Clinical practice guidelines; Reminders; In-service training; Continuous education; Supervision</td>
<td>Clinical incident reporting; Clinical practice guidelines in PHC facilities; Local opinion leaders; Continuous quality improvement (including lean management).</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Accountability arrangements</td>
<td>Audit and feedback</td>
<td>Public release of performance information; Social accountability</td>
<td></td>
</tr>
<tr>
<td>Financial arrangements</td>
<td>Performance-based financing (Includes supply-side Results-Based Financing, Pay for Performance, and other provider incentives and rewards)</td>
<td>Performance-based financing (Includes supply-side Results-Based Financing, Pay for Performance, and other facility-based incentives and rewards)</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>
although there may be supplemental investment financing of some inputs, including training and equipment to enhance capacity or quality; and they are not made unless and until results or performance are satisfactory; and,

- **Performance-based financing** is a form of RBF distinguished by three conditions. Incentives are directed only to providers, not beneficiaries; awards are purely financial—payment is by fee for service for specified services; and payment depends explicitly on the degree to which services are of approved quality, as specified by protocols for processes or outcomes;

- **Pay-for performance, performance-based payment and performance-based incentives** can all be considered synonyms for RBF. Performance in these labels means the same thing as results, and payment means the same thing as financing.

### Outcomes included in the evidence gap map

Outcomes were categorized following the guidelines developed for EPOC systematic reviews and adapted for the identification of gaps in research on performance management in primary care systems in LMIC. Relevant outcomes are, therefore, those that can be actionable for the intended users: research groups, funding agencies, and performance measurement and management practitioners in primary care systems in LMIC. Based on these considerations outcomes of interest will be wide in scope; can occur across short- and long-term timeframes; can be observable at various levels within a system (individual, organizational, social); and, can include desirable as well as undesirable, adverse effects. Equity effects constitute a unique, cross-cutting category, the inclusion of which is recommended in evidence gap maps. The priority-setting exercise that will follow the completion of this EGM may result in the identification of primary and secondary outcomes; at this stage, however, the study aims to scope the largest number of relevant outcomes within available operational constraints.

The main categories included in this EGM are listed in Table 2. They include: 1) provider and managerial outputs and outcomes, defined as individual, provider and managerial staff effects, and exemplified by changes in workload, work morale, stress, burnout, sick leave, and staff turnover; 2) patient outcomes, defined as changes in health status or on patient health behaviors; 3) organizational outcomes, defined as organizational-level effects within and across facilities and networks of primary care such as quality of care process improvements, patient satisfaction, perceived quality of care, workforce retention, organizational culture, and unintended outcomes (gaming, shirking, shaming, data falsification, etc.); 4) population-level outputs and outcomes, defined as aggregate, health and equity effects accruing defined populations, including utilization of specific primary care services (for instance, 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; and, 5) social outcomes defined as non-health, social, economic, or cultural effects affecting defined populations, such as changes in community participation, non-health equity effects, non-health adverse effects or harm, and other unintended social outcomes. Table 2 lists each outcome category and provides examples of specific types of results within each category.

### Table 2. Outcomes of relevance to users of the evidence gap map.

<table>
<thead>
<tr>
<th>Provider and managerial outputs and outcomes</th>
<th>Patient outcomes</th>
<th>Organizational outcomes</th>
<th>Population health outputs and outcomes</th>
<th>Social outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workload</td>
<td>Health status outcomes: a) Physical health and treatment outcomes such as mortality, and morbidity; b) Psychological health and wellbeing; c) Psychosocial outcomes such as quality of life, social activities</td>
<td>Quality of care process improvements; Adherence to recommended practice or guidelines; Patient satisfaction Perceived quality of care Workforce retention Changes in organizational culture Unintended organizational outcomes</td>
<td>Utilization of specific services (example: number of antenatal visits) Coverage of specific services or interventions (example: proportion of pregnant women receiving antenatal care; proportion of pregnant women delivering in facilities; coverage rate of specific vaccines) Access to primary care services (example waiting times) Health equity effects Adverse health effects or harm Unintended population outcomes</td>
<td>Community participation Other equity effects Unintended social outcomes</td>
</tr>
<tr>
<td>Work morale</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stress</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burnout</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sick leave</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staff turnover</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Adapted from: Cochrane Effective Practice and Organization of Care (EPOC). What outcomes should be reported in EPOC reviews? EPOC resources for review authors, 2017.
How performance measurement and management may work in primary care delivery systems

Based on the above considerations, the performance measurement and management framework developed for this study is delineated in Figure 2.

The framework highlights some of the dynamic and complex relationships between interventions and outcomes and characterizes a process of multi-level change in a primary care delivery system. The process of change described in this framework adheres to the following logic:

- Performance management interventions operating at individual, organizational, and social levels can initially trigger short-term changes in healthcare as well as in managerial and ancillary service process improvements. Such changes would be the result of short cycles of experimentation with technological, managerial and clinical innovations.\(^{80,81}\)

- The repetition of these cycles through time, and the utilization of the information derived from performance measurement by system actors would lead to the generation of proximal outputs such as policies, regulations, clinical guidelines, and standard operating procedures, as well as negative or unintended outcomes.

- If effectively implemented, these new routines and processes would lead to intended and unintended proximal outcomes including changes in the behaviors of healthcare providers, primary care managers, and policy-makers. Proximal health effects could include the adoption of improved clinical behaviors by providers; quality improvement and safety outcomes at the patient- and facility-levels; increased service utilization and effective coverage; positive and negative equity effects; and, adverse or unintended effects. Examples of negative or unintended effects of performance management interventions have been reported in the literature, including gaming, shirking and cream-skimming.\(^{82-85}\)

- If sustained through time and effectively implemented, additional desirable outcomes from iterative cycles of innovation, measurement and improvement may include increased retention of the workforce; increased productivity and efficiency; or improved equity, among others.

- Continuous cycles of performance measurement and management would also lead to the emergence or reinforcement of organizational-level capabilities and resources that could sustain performance improvements at higher-levels within the primary care system leading to, in some but not all instances, reinforcing cycles of improvement and organizational learning.

- The reiteration of these reinforcing cycles would be necessary conditions for the sustained generation of organizational-level level outcomes such as improved quality, patient safety, customer satisfaction and, distantly, for the generation of population-level health outcomes. Private sector organizations outcomes may include profits, market share, efficiency and productivity gains, and customer satisfaction, among others, but these have been excluded from this study.

The framework contains three additional elements that would generate interdependence and non-linearities in the behavior of a primary care performance measurement and management system and that would help explain how performance measurement and management systems could work or not, and why. These include 1) the recursive linkages among system elements, described in Figure 1 as bidirectional arrows which will likely generate feedback effects; 2) the dynamic interaction between context and system actors, which will likely introduce context-specific variations in the outcomes from performance management interventions; and, 3) the repetition of performance measurement and management cycles as a precondition for the generation of sustained change/improvement through time.

Also, given the well-known limits to the adoption and use of evidence by healthcare system actors at all levels,\(^{86-91}\)

---

**Figure 2.** Performance measurement and management framework in primary care delivery systems.
the use of performance information is a critical, intermediate factor in the process of production of downstream outcomes. In ways that are similar to how the results of evaluation studies may or may not be used, system actors’ use of performance information is oftentimes implicit in assumptions about how performance measurement and management systems are supposed to generate multi-level outcomes. Performance information use can be defined as “the assessments, decisions, or attitudes that primary care system actors and stakeholders hold towards the interventions that are the object of the PM system”27.

Primary care system actors’ assessments, decisions or attitudes can be triggered or not in response to 1) the performance measurement and management interventions in use; and, 2) the contextual conditions in which they are embedded11,98–102. For performance effects, the supply of performance information needs to be accompanied by individual and organizational decisions to act upon it. Unfortunately, production of the former does not always guarantee the achievement of the latter40,103. Also, the assumption that adopting performance information will only have positive effects has also been proven not to be correct at all times104–106.

Context factors, or the environment or setting in which the proposed process of change is to be implemented can exert influence through interactions that occur at multiple levels (individual, interpersonal, organizational, community and societal) within the primary care system. Such factors can facilitate or inhibit the effects of performance measurement and management systems and are exemplified by the composition and dynamics of the institutional primary care setting (policies, legislation, and sector-specific reforms, among others); the degree of autonomy or flexibility granted to primary care delivery actors to innovate and implement organizational changes; and by social and political pressures for transparency, accountability or social control, among others. System antecedents, such as experiences with previous institutional reforms, and the readiness for change in the primary care system, have also been shown to have effects on the acceptance and assimilation of performance improvements107,108. Finally, ancillary components like technical assistance, monitoring and evaluation, and training, among others, should also be considered as relevant factors that can contribute or create obstacles in the generation of performance improvements25,7,62. Given operational limitations, contextual factors and ancillary components will not be included in the scope for this EGM.

Study objectives
This evidence gap map aims to identify and describe the existing evidence on the effects of interventions in the area of performance measurement and management in primary care delivery systems in LMICs. Also, to identify evidence gaps where new primary studies or systematic reviews could add value and provide easy access to the best available existing evidence on intervention effects in this area. The resulting EGM will inform the development of a prioritized research agenda for primary care delivery systems in LMICs.

Methods
Overall approach
The team will follow the methodology to produce evidence gap maps developed by 3ie109,110. The methodology was developed as a tool to systematically map evidence and research gaps on intervention effects for a broad topic area. In doing so, EGMs can help inform strategic use of resources for new research by identifying ‘absolute gaps’ where there are few or no available impact evaluations, and ‘synthesis gaps’ where there are clusters of impact evaluations but no available high-quality systematic reviews. By making existing studies easily available to researchers and describing the broad characteristics of the evidence base, the EGM can also inform the methods and design of future studies. Evidence gap maps can also facilitate the use of evidence to inform decisions by providing collections of systematic reviews that are critically appraised and ready for use by various decision makers. The methods used to develop EGMs are informed by systematic approaches to evidence synthesis and review and include key characteristics such as explicit inclusion/exclusion criteria and a systematic and transparent approach to study identification, data extraction and analysis. We describe our methods in more detail below.

Criteria for including and excluding studies
The process starts with developing the scope for the EGM by creating an evidence-informed framework that serves for the identification of the interventions and outcomes that are relevant for the domains under study.

Table 1 and Table 2 define the final intervention and outcome inclusion criteria. To be included, studies have to assess the effect of at least one of these interventions on one of the outcomes.

On the supply-side of performance management, the study will exclude interventions focusing on the delivery of vector-control and epidemiological surveillance activities; and services delivered in hospitals. Demand-side interventions will be excluded from the study (financial interventions such as conditional cash transfers; communication for behavior-change interventions; and social marketing interventions, among others).

Types of included study designs
We will include studies designed to assess the effects of interventions, and systematic reviews of such studies, as defined below:

1) Explicitly described as systematic reviews and reviews that describe methods used for search, data collection and synthesis as per the protocol for the 3ie database of systematic reviews (Snilstveit et al., 2018).

2) Impact evaluations, defined as program evaluations or field experiments that use experimental or observational data to measure the effect of a program relative to a counterfactual representing what would have happened to the same group in the absence of the program. Specifically we will include the following impact evaluation designs: Randomized controlled trials (RCT) where the
intervention is randomly allocated at the individual or cluster level; Regression discontinuity design (RDD); Controlled before and after studies using appropriate methods to control for selection bias and confounding such as Propensity Score Matching (PSM) or other matching methods; Instrumental Variables Estimation or other methods using an instrumental variable such as the Heckman Two Step approach; Difference-in-Differences (DD) or a fixed- or random-effects model with an interaction term between time and intervention for baseline and follow-up observations; Cross-sectional or panel studies with an intervention and comparison group using methods to control for selection bias and confounding as described above; and, Interrupted-time series (ITS), a type of study that uses observations at multiple time points before and after an intervention (the ‘interruption’). We will only include ITS studies that use at least three observations before and three observations after the intervention.

Efficacy trials and systematic reviews of efficacy trials will be excluded. Broadly, efficacy trials determine whether an intervention produces the expected result under ideal/controlled circumstances, whereas effectiveness trials measure the degree of beneficial effect under “real world” settings. However, the distinction between these two types of studies is generally considered as a continuum rather than a clear dichotomy and in practice it can be difficult to clearly categorize a trial as either effectiveness or efficacy\(^{14}\). We will therefore draw on the criteria developed by Snilstveit et al. to aid the identification of efficacy trials for exclusion from the EGM. The adapted criteria are as follows:

- **Research Objective:** Is the study primarily designed to determine to what extent a specific technique, technology, treatment, procedure or service works under ideal condition rather than attempt to answer a question relevant to the roll-out of a large program?

- **Providers:** Is the intervention primarily delivered by the research study team rather than primary health care personal/trained laypersons who don’t have extensive expertise?

- **Delivery of intervention:** Is the intervention delivered with high degree of assurance of delivery of the treatment? (Is the delivery tightly monitored/supervised by the researcher following specific protocols; Is adherence to the treatment monitored closely with frequent follow-ups?)

Other inclusion and exclusion criteria

In addition, studies have to be conducted in a low- or middle-income country based on a search template from Cochrane EPOC Group with LMICs defined by World Bank. Moreover, studies have to be published in English in the year 2000 and after. We will include studies regardless of status (ongoing or completed) and type of publication, published (e.g. journal article, book chapter) and unpublished (e.g. report or working paper). Studies published before 2000 will be excluded. Finally, we will also exclude studies exclusively focused on high-income countries, or systematic reviews focusing on a single country.

**Search strategy**

We have developed a systematic search strategy in collaboration with two information specialists. We developed a detailed search string for searching bibliographic databases and relevant portals. A sample strategy was developed for Medline, (see Supplementary File 1) and covers a detailed explanation of the search terms used based on an initial set of English search terms relevant to the main concepts of our inclusion criteria, including intervention, study design and population (low- and middle-income countries). These were combined using appropriate Boolean operators. All search strategies used in the study will be published along with study results.

We will identify potential studies using three strategies as listed below:

- Advanced search of the following bibliographic databases such as Medline (Ovid), Embase (Ovid), CAB Global Health (Ovid), CINAHL (Ebsco), Cochrane Library, Scopus (Elsevier), and Econlit (Ovid);

- Search of key institutional databases, repositories of impact evaluations and systematic reviews and other sources of grey literature such as the International Initiative for Impact Evaluation Impact Evaluation and Systematic Review repositories; Cochrane Effective Practice and Organization of Care (EPOC); the Abdul Latif Jameel Poverty Action Lab (J-PAL); The World Bank’s Independent Evaluation Group; Inter-American Development Bank repository; and, American Economic Association Register;

- Snowballing the references in appraised systematic reviews and citation tracking of included studies using Scopus and contacting authors, when required.

**Procedures for screening and data extraction**

Following the search, we will import all records into EPPI reviewer 4. Following the removal of duplicates, we will combine manual screening and text mining to assess studies for inclusion at the title and abstract stage. To ensure consistent application of screening criteria for all screeners, we plan to assess the same random sample of 100 abstracts. Any discrepancies will be discussed within the team and inclusion criteria will be clarified if necessary. Following this initial set of 100, we will move to single screening with “safety approach”, whereby there is an option to mark unclear studies for review by a second reviewer\(^{15}\).

Once all screeners have been trained, we will screen a random sample of 500 abstracts to train EPPI reviewer’s priority screening function. The priority screening function can be used at the title/abstract screening stage to prioritize the items most likely to be ‘includes’ based on previously included documents. Using priority screening in this way allows for the identification of inculdable records at an earlier stage in the review process.
so that work can begin earlier on full-text screening and data extraction.

Depending on the number of search hits, we may also make use of EPPI reviewer’s auto-exclude function to auto-exclude studies from the search that have less than a ten per cent probability score of inclusion. This function classifies un-screened studies into ten percent intervals of probability of inclusion, based on keywords included in previously included and excluded studies.

Because of time and resource constraints we will not conduct independent double screening of all studies that will be considered at full text. To minimize bias and human error we will however double screen a sample of studies at the beginning to ensure inter-rater reliability between screeners. In addition, we will take a “safety first” approach as at the title and abstract stage, whereby any studies where the first screener is uncertain about inclusion/exclusion will be screened by a second person. All studies identified for inclusion will be effectively screened by a second/third person during data extraction.

We will use a standardized data extraction form in Microsoft Excel to systematically extract data from all included studies, including bibliographic details, intervention type and description, outcome type and definition, study design, and geographical location. We will also assess the extent to which studies incorporate equity, and extract information about if and how studies consider vulnerable and marginalized groups. To do so we will draw on the PROGRESS-Plus framework which outlines dimensions that may give rise to inequity in either access to services, or final health outcomes. In particular, we will consider the following dimensions: Place of residence (location of household such as distance from health facility, or rural/urban), ethnicity, culture and language, gender, socioeconomic status and other vulnerable groups (open category to be used iteratively to record details of any vulnerable groups identified during coding).

For each study we will assess if they consider equity for any of these dimensions, and if so how, giving the following options:
1) Contains equity-sensitive analytical frameworks/theory of change; 2) Uses equity-sensitive research questions; 3) Follows equity-sensitive methodologies (sub-group analysis); 4) Contains equity-sensitive methodologies: additional study components to assess how and why (including mixed and qualitative methods); 5) Uses any other methodology that is equity sensitive that is not covered by the other options; 6) Uses equity-informed research processes (who are the respondents, who collects data, when, where etc.); 7) Addresses interventions targeting specific vulnerable groups - Looks at the impact of an intervention that targets specific population groups; and/or; 8) Measures effects on an inequality outcome.

For multi-arm trials testing different interventions, each comparison arm will be treated as an individual study for the coding of interventions. We will report both number of studies and number of papers identified. In addition, we will report on the number of linked studies. Studies will be considered linked if there are multiple papers by the same study team on the same impact evaluation reporting different outcomes or different follow-up periods. If they report the same information, the study will be excluded as a duplicate.

A full list of descriptive data to be extracted is included in the coding tools in Supplementary File 2. We will begin the coding process with a training with the whole research team. This training will involve coding one included systematic review and one included impact evaluation as a group to familiarise all coders with the coding tools. The entire research team will then independently pilot the coding tool on the same small subset of studies to ensure consistency in coding and to resolve any issues or ambiguities. We will start this process with two systematic reviews and two impact evaluations, and test an additional small subset if issues or discrepancies remain in the application of the tool. Data extraction will then be completed by a single coder.

To minimize bias and human error we will however review the data extraction of a sample of studies.

We will follow the adapted SURE checklist, available in the 3ie systematic review database protocol for appraisal of systematic review. This checklist is based on the SURE Collaboration checklist for deciding how much confidence to place in the findings of a systematic review, giving systematic reviews a rating of high, medium or low confidence (Supplementary File 3). We will produce a user friendly summary for all reviews of high confidence. All systematic reviews will be appraised by at least two people, and shared with study authors before publication.

**Statistical analysis plan and EGM visualisation**

Upon completion of the data collection, findings will be initially presented in a visual interactive format using 3ie’s custom built platform and accompanied by a detailed report.

The visual, online EGM will be build by, first, transforming the intervention-outcome framework into a matrix followed by the uploading of cvs files with data for all the studies included in the map (intervention, outcome, study type, impact evaluation study design, systematic review confidence level, geographical location and equity focus). This data will automatically populate the framework matrix to indicate the relative availability of evidence. This will be used to identify and describe absolute evidence gaps (no studies) and synthesis gaps (sizeable impact evaluation literature, but no high confidence SR). In addition the map will contain descriptions of the characteristics of the evidence base by using graphs, figures and descriptive statistics.

**Discussion**

Evidence gap maps consolidate what is known about what works in specific development sectors or thematic areas. They provide thematic collections of evidence from systematic reviews and impact evaluations structured around a framework of interventions and outcomes of relevance to any given sector. They also provide a graphical display of the volume of impact evaluations and/or quality of existing systematic reviews. EGM can be used to
identify areas where there is need for the generation of new or more rigorous research evidence; also, to inform decisions by policymakers and development practitioners as policies and programs are designed.

Current best-practice in the design of evidence gap maps recommends that EGMs have a pre-specified protocol, have a systematic search strategy, contain precise and clear criteria for inclusion and exclusion, and systematically report all eligible studies. This protocol is the first step in ensuring compliance with such practices. As the visual interface is built, a final report will be submitted for peer-reviewed publication and will include a summary of the findings from the evidence gap analysis and recommendations for a prioritized research agenda on performance measurement and management in LMIC.

This evidence gap map aims to identify and describe the existing evidence on the effects of interventions in the area of performance measurement and management in primary care delivery systems in LMICs and, also, to provide easy access to the best available existing evidence on intervention effects in this area. As a result, the EGM will inform the development of a prioritized research agenda for primary care delivery systems in LMICs.

Dissemination of findings
Findings will be presented at the 5th Global Health System Symposium in October, 2017 after which a paper will be submitted for peer-reviewed publication. We will also publish the results in the form of an interactive Evidence Gap Map, which will be made freely available from the 3ie website. We will use our institutional channels to disseminate our findings as widely as possible, including via our websites, social media platforms and events beyond the Global Health System Symposium.

Study status
By the time of submission of this paper, the framework, search strategies and data extraction tools included in this protocol have been completed. Data collection, analysis and development of the graphical interface will be completed by September 2018. A paper summarizing study results and implications will be submitted by December, 2018.

Data availability
No data is associated with this article

Competing interest
No competing interests were disclosed.

Grant information
Bill and Melinda Gates Foundation [OPP1154415, OPP1149078].

This work was supported by the Bill & Melinda Gates Foundation through grants to George Washington University [OPP1154415], and Ariadne Labs through Brigham and Women's Hospital [OPP1149078].

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 - Coding tools containing full list of descriptive data to be extracted
Click here to access the data.

Supplementary File 3 – Systematic Reviews critical appraisal checklist
Click here to access the data.

References


Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 02 July 2018

https://doi.org/10.21956/gatesopenres.13899.r26534

© 2018 Eliakimu E. This is an open access peer review report distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Eliudi Eliakimu
Health Quality Assurance, Ministry of Health and Social Welfare, Dar es Salaam, Tanzania

As all countries of the World work towards attainment of the Sustainable Development Goals (SDGs) including SDG-3: “Ensure healthy lives and promote well-being for all at all ages”, the need for a stronger Primary Health Care (PHC) System is evident. The need for a well-functioning PHC systems particularly in Low and Middle-Income Countries (LMICs) is critical given the fact that stronger PHC system in LMICs is essential for achieving the Universal Health Coverage (UHC) target. The purpose of this referee report is to review the article on “Evidence Gap Map of Performance Measurement and Management in Primary Health Care Delivery Systems in Low and Middle-Income Countries…..” by Munar et al. to examine whether it is scientifically sound. The referee report examine the article based on the following parts as they appear in the paper: background; interventions and outcomes of interest; how performance measurement and management may work in primary care systems; study objectives; methods; discussion; dissemination of findings; supplementary materials; and references.

Background – The authors have presented well the background information and indeed they have done good job by bringing in an interdisciplinary perspective in the performance measurement and management in PHC in LMICs. Their generic framework presented in figure 1 is comprehensive enough and it clearly shows the interplay between the performance measurement and management steps and the context in health sector which encompass both institutional and organizational contexts as well as the context outside the health sector. This brings a robust perspective on what constitutes performance of PHC system in LMICs.

Interventions and outcomes of interest – The authors have clearly defined the key concepts involved and description of the interventions and outcomes of interest. This part is well presented, however the authors need to work on the following minor issues to improve this part. First, table 1 in page 5 of 14 contains a total of 12 interventions (i. Clinical practice guidelines; ii. reminders; iii. in-service training; iv. continuous education; v. supervision; vi. clinical incident reporting; vii. local opinion leaders; viii. continuous quality improvement [including lean]; ix. performance based financing; x. public release of performance information; xi. social accountability; and xii. audit and feedback), which are spread in the three categories of interventions, but the authors have made an error by indicating that in total there are 11 interventions. Secondly, they need to correct spelling error for the word “accountability” in right column.
middle row where it refers to social accountability. Also, on the same page 5 of 14, in the paragraph with subheading “accountability arrangements”, the authors need to check the spellings for “intervention” in the fourth line in that paragraph since it is not clear whether they wanted to say “such intervention” or “such interventions”. The outcomes included in Table 2 on page 6 of 14 in the column named as “provider and managerial outputs and outcomes” are appropriate, however, I suggest to the authors to consider adding provider availability at work (i.e., looking at absenteeism of health workers in health facilities). This is an important management element in health facilities in LMICs. For example, in Tanzania, absenteeism has been documented to be 14% and in other LMICs it has been documented to be in a range of 14%—60%.

Some countries are trying to address this challenge, for example, as a way of improving performance management in PHC facilities in Tanzania through an initiative referred to as Big Results Now (BRN), the delivery bureau through stakeholders consultations in a setting contextualized as “Laboratory” they identified use of timesheet as one of measures to address absenteeism. Therefore, I strongly recommend to the authors to consider this as part of the managerial outputs and outcomes. Also, on page 6 of 14 in the paragraph starting with “The main categories included in this EGM…….” The fourth category “4) population-level……. equity effects accruing defined population……,” I suggest to add the word “in” between accruing and defined.

How performance measurement and management may work in primary care systems – This part is well presented and figure 2 that depicts the authors’ framework for performance measurement and management in PHC systems is very comprehensive with clear linkages between its components.

Study objectives – The objectives of the study are well presented and succinct. I am convinced that upon realization of the objectives, the outputs of the study will synergize very well with the work done so far (as well as ongoing work) in the Primary Health Care Performance Initiative (PHCPI).

Methods – The authors have presented a detailed methodology which is well organized narrating the inclusion and exclusion criteria; the types of study designs that are included; search strategy; screening procedure; and statistical analysis and EGM visualization. To the best of my knowledge this part is correct. There is a very minor correction on page 9 of 14 under the subheading “procedures for screening and data extraction”, specifically in the second paragraph (fifth line), they need to change the word “includes” to read as “included”, hence the sentence will read: ……..”most likely to be included based on …. “

Discussion – It is brief and clear that orients a reader on what is an EGM and its value in decision making which is in line with the existing literature. However, I suggest that the authors need to add a bit of information on how to read the EGM so that readers (including decision makers) who are new to the concept can easily follow and understand the EGM when it is released and disseminated.

Dissemination of findings – Excellent explanations on how to disseminate the findings including the use of social media. However, the authors need to correct the year for the 5th Global Health System Symposium. I hope that they mean “the 5th Global Health System Symposium in October 2018 and not 2017” as indicated in the paper on page 11 of 14.

Supplementary materials – The supplementary materials are well presented and comprehensive. The search strategy (File 1) is comprehensive. The coding tools (File 2) is also comprehensive and well presented, however, on page 2 and 6 in the middle column under the heading “population health outputs and outcomes (defined as………..effects accruing defined populations)”, there is a need to add “in” between accruing and defined. File 3 contains a very comprehensive checklist.
References – The reference list is correctly presented.

References
1. van Weel C, Kidd MR: Why strengthening primary health care is essential to achieving universal health coverage. CMAJ. 2018; 190 (15): E463-E466 PubMed Abstract I Publisher Full Text

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

Author Response 02 Jul 2018

Wolfgang Munar, George Washington University, Washington, USA

Dear Prof. Eliakimu: On behalf of the authors, we all appreciate your detailed comments very much. We will make sure to edit and update the paper to reflect them. The well-founded suggestion regarding absenteeism is quite relevant. The team will discuss and make a decision about its inclusion in the evidence gap map.

Wolfgang Munar

Competing Interests: No competing interests were disclosed.

Reviewer Report 25 June 2018
The research is pertinent due to the lack of evidence about primary health care performance in low- and middle-income countries. The objective is clear however the goal of this specific article is to describe the gap map methodology as a first step to develop the gap map.

I recommended to differentiate between low- and middle-income countries because they have very different context and capacity to produce and publish evidence on primary health care.

The use of the International Initiative for Impact Evaluation allows the researchers to compare the results with other studies and to have a standard methodology.

The performance management system included implementation strategies and financial arrangements. However, the methodology excluded demand-side interventions (financial).

The performance framework in primary care delivery systems is already an important product of this research including evaluation cycle and context. However, the link between the two is not clear, the two directions arrow is not enough to explain the interaction between the context and the outcomes.

The framework is an excellent exercise combining interventions, outputs and outcomes. But, the context is not present in this framework. Some indicators need to be more precise, for example the “number of antenatal visits” (% of pregnant women with more than 2, 3 or 4 antenatal visits?)

The use of grey literature is an important aspect of this project because most of the evaluations on primary health care intervention has not been published in peer journals. I recommended to add web pages from Ministries of Health and Bireme because they have a data base on grey literature. The research will review only articles in English, I recommend to included French and Spanish to increase the countries and articles from Latin America and some African countries.

The discussion not included the limitations of this study, related for example to the limited quality of data in low- and middle-income countries.

They mention that the “Findings will be presented at the 5th Global Health System Symposium in October 2017” but the article is not published until now.

The bibliography is extensive and updated and show a big effort in terms of analysis of existing evidence.

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 systems Health provision in rural settings.

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

---

**Author Response 02 Jul 2018**

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

Dear Prof. Bernal-Acevedo- thanks a lot for your comments. They are all relevant and applicable. Your comment about the generic and possibly inappropriate aggregation of low- and middle-income countries is one that we expect to address in the findings as it may well require definitions that differentiate institutional and organizational contexts in addition to, the conventional income-based definitions used in the "LMIC" construct.

Wolfgang Munar on behalf of the authors

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