

REVIEW

Evaluating health systems strengthening interventions in low-income and middle-income countries: are we asking the right questions?

Taghreed Adam,^{1*} Justine Hsu,² Don de Savigny,³ John N Lavis,⁴ John-Arne Røttingen⁵ and Sara Bennett⁶

¹Alliance for Health Policy and Systems Research, World Health Organization, Geneva, Switzerland, ²London School of Hygiene and Tropical Medicine, London, UK, ³Swiss Tropical and Public Health Institute and University of Basel, Basel, Switzerland, ⁴McMaster Health Forum, Centre for Health Economics and Policy Analysis, Department of Clinical Epidemiology and Biostatistics, and Department of Political Science, McMaster University, Hamilton, Canada, ⁵Harvard Kennedy School, Cambridge, MA, USA and Institute for Health and Society, University of Oslo, Oslo, Norway and ⁶Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

*Corresponding author. Alliance for Health Policy and Systems Research, World Health Organization, 1211 Geneva 27, Switzerland. Tel: +41-22-7913487. E-mail: adamt@who.int.

Accepted 17 August 2012

In recent years, there have been several calls for rigorous health policy and systems research to inform efforts to strengthen health systems (HS) in low- and middle-income countries (LMICs), including the use of systems thinking concepts in designing and evaluating HS strengthening interventions. The objectives of this paper are to assess recent evaluations of HS strengthening interventions to examine the extent to which they ask a broader set of questions, and provide an appropriately comprehensive assessment of the effects of these interventions across the health system. A review of evaluations conducted in 2009–10 was performed to answer these questions.

Out of 106 evaluations, less than half (43%) asked broad research questions to allow for a comprehensive assessment of the intervention's effects across multiple HS building blocks. Only half of the evaluations referred to a conceptual framework to guide their impact assessment. Overall, 24% and 9% conducted process and context evaluations, respectively, to answer the question of whether the intervention worked as intended, and if so, for whom, and under what circumstances. Almost half of the evaluations considered HS impact on one building block, while most interventions were complex targeting two or more building blocks. None incorporated evaluation designs that took into account the characteristics of complex adaptive systems such as non-linearity of effects or interactions between the HS building blocks.

While we do not argue that all evaluations should be comprehensive, there is a need for more comprehensive evaluations of the wider range of the intervention's effects, when appropriate. Our findings suggest that the full range of barriers to more comprehensive evaluations need to be examined and, where appropriate, addressed. Possible barriers may include limited capacity, lack of funding, inadequate time frames, lack of demand from both researchers and research funders, or difficulties in undertaking this type of evaluation.

Keywords Evaluation, health systems, methods, systems thinking, health policy and health systems research

KEY MESSAGES

- Despite calls for rigorous evaluations of efforts to strengthen health systems, very few evaluations used a comprehensive and holistic approach to assess the impact of interventions with system-wide effects.
- Our findings suggest the need for better understanding the barriers towards taking a more holistic evaluation approach, when evaluating interventions with system-wide effects.
- Possible barriers to more comprehensive evaluations may include limited capacity, lack of funding, inadequate time frames, lack of demand from both researchers and research funders, or difficulties in undertaking this type of evaluation.
- There are several untapped resources that can contribute to more comprehensive evaluations, including systems thinking concepts, tools and approaches; and perspectives and approaches used in relevant disciplines, such as social sciences and health policy analysis.

Introduction

It is now well accepted that strong health systems are paramount to achieve health systems goals (Evans *et al.* 2008). Consequently, several new interventions or initiatives have been launched, at global and national levels, to address some of the bottlenecks to scale up essential health interventions and to strengthen some components of the health system (van Etten *et al.* 2006). At the global level, the GAVI Alliance, the Global Fund for AIDS, Tuberculosis and Malaria, and other major funders have been explicitly encouraging the inclusion of health systems strengthening interventions in grant applications in recent years, and several international initiatives dedicated to strengthening health systems have been established, e.g. the Implementation Research Platform hosted by the Alliance for Health Policy and Systems Research, the International Health Partnership (IHP+) and the High-level Taskforce on Innovative Financing for Health Systems (Bennett *et al.* 2008; Alliance for Health Policy and Systems Research 2011).

At the same time, there has been increasing recognition of the need for more rigour in designing and evaluating the effects of global health initiatives and interventions that aim, at least in part, to strengthen health systems in order to improve the population's health (Evans *et al.* 2008; Mills *et al.* 2008; de Savigny and Adam 2009; Swanson *et al.* 2009).

Acknowledging this need, several recent publications have sought to define health systems and their boundaries (World Health Organization 2007), what is meant by health systems strengthening (World Health Organization 2007; Swanson *et al.* 2010), the nature of health systems research (Remme *et al.* 2010; Bennett *et al.* 2011; Mills 2012), and the importance of systems thinking in designing, implementing and evaluating health systems strengthening interventions (Leischow *et al.* 2008; Shiell *et al.* 2008; de Savigny and Adam 2009).

In a rapidly evolving field and amid continuing calls for more rigorous health policy and health systems research (World Health Organization 2005; Mills *et al.* 2008; Swanson *et al.* 2009; Bennett *et al.* 2011), what has been the practice in the field of health system evaluations in recent years and how well do such evaluations address contemporary issues and recommendations? More specifically, our recent publication on systems thinking and its value for the design, implementation and evaluation of health systems strengthening interventions argued for the need for a more comprehensive and systematic

approach in thinking through the underlying causes of health systems problems, and in designing new interventions and their evaluations (de Savigny and Adam 2009). In our background review of the peer-reviewed and grey literature up to 2008, we found a very limited number of evaluations assessing the wider impact of complex health interventions on health systems. When they did, the evaluations were conducted in high-income countries and on health-outcome oriented interventions such as tobacco control, obesity or cancer (Best *et al.* 2007; Butland *et al.* 2007). This paper seeks to understand what has happened since then. It does so through a review of the peer-reviewed and grey literature to map out current practices in evaluating health systems strengthening interventions in low- and middle-income countries (LMICs) and to assess the extent to which health systems effects have been explored.

For the purpose of this review, health systems strengthening (HSS) interventions include those system-level interventions that are directly targeting one or more of the six health system building blocks and their sub-components as defined by the World Health Organization (WHO) (World Health Organization 2007); or disease-specific interventions or programmes that also have important system-wide effects, e.g. scale-up of antiretroviral therapy for HIV/AIDS (de Savigny and Adam 2009). This approach implicitly reflects the relationship between the different health system components as well as the interests and power of its different actors and beneficiaries (both supply and demand side).

The overall objectives of this paper are to assess the scope and research questions explored in recent HSS evaluations. More specifically, to assess whether the research questions attempted to explore the intervention's effects across multiple health system building blocks and actors, and if they did, to what extent, and with what methodological approaches.

The objective is not to appraise the quality of evidence, e.g. whether the evaluation used appropriate study design or methods. It is rather to assess whether they ask a broader set of questions relevant for policy making. For example, whether or not the intervention worked as intended; and what are the elements that contributed to its success or failure, which could influence the replication of its impact in other settings. If the intervention has broader implications across the health system, to assess what these are with respect to both intended and unintended effects (Leischow and Milstein 2006; Trochim *et al.* 2006; de Savigny and Adam 2009; Paina and Peters 2011).

Without such deep analysis of the process and context around which the intervention worked, and, as relevant, its broader effects on the health system as a whole, evaluations may over- or under-estimate the actual impact of the intervention, or overlook important effects on the system itself or other interventions already in place (Rychetnik *et al.* 2002; Savedoff *et al.* 2006).

Methods

Literature search

We conducted a systematic search in Medline and Embase as well as the individual websites of 36 institutions including research funders, think tanks, academic and research institutions, and partnerships and alliances identified through a web-based search or known to conduct and publish evaluations of public health interventions.

The search strategy builds on previous literature reviews with similar objectives (Bennett *et al.* 2008; Lewin *et al.* 2008; Ritz *et al.* 2010; Adam *et al.* 2011), with further refinements and iterative testing of individual search terms (see Supplementary Data Web Annex 1). Articles were included if they met the following criteria:

- (a) *Evaluation studies* defined as studies that report on the output, outcome or impact of an intervention on the health system or studies assessing if and how a programme worked.
- (b) *Health system strengthening intervention* defined as ‘system-level’ interventions directly targeting one or more of the six health system building blocks and their sub-components (see Table 4); or disease-specific interventions expected to have large system-wide effects (de Savigny and Adam 2009).
- (c) *Low- and middle-income countries* based on the World Bank classification (World Bank 2011).

For the second criteria above, we sought to assess the general relevance of interventions and to exclude those that would only weakly impact the performance of health systems, or the values and interests of its actors or beneficiaries. Our common interpretation considered this to involve:

- Interventions with system-level changes as opposed to changes at the organizational level (e.g. interventions involving changes to patient access to care have system-level repercussions and were included, but not those focusing on modifications of patient flow within a health facility, which is unlikely to have system-wide impact).
- A need for a systems approach or complex interventions that require identifying and evaluating interactions between health system building blocks/sub-systems (e.g. a systems approach for evaluating training aimed at improving the quality of care provided, but not for evaluating the relevance of the training material).
- Evaluations of the cost-effectiveness of system-level interventions, but not simple costing analyses (e.g. cost-effectiveness analyses of task-shifting, but not a costing study of malaria case management).

Given our prior publication on systems thinking and evaluation (de Savigny and Adam 2009), the search focused on articles published in 2009–10 in order to assess the most recent practice in the field since the prior publication, and to get a good picture of the most recent evaluations. There was no language restriction.

Literature screening

Two independent raters (JH and TA) screened the first 100 articles against the inclusion criteria to determine inter-rater agreement. Any disagreement in article selection was discussed until consensus was reached. We then calculated the level of inter-rater agreement using a simple Kappa analysis (Cohen 1960); at least substantial agreement (i.e. kappa exceeding 0.6) was desired for a decision to continue with a single rater (Landis and Koch 1977). The calculated kappa score was 0.81, classifying the level of agreement to ‘almost perfect’ (Landis and Koch 1977). We therefore continued the screening for article selection with one rater (JH).

Data abstraction and analysis

We retrieved the full text of all articles that met the inclusion criteria. Data abstraction included the following variables: country where the evaluation was conducted; type of intervention, i.e. system-level or disease-specific (see definition above); name and brief description of the intervention; primary health system building blocks targeted by system-level interventions, as defined in the studies (see Table 4 and Supplementary Data Web Annex 2 for a description); whether the scope of the evaluation was narrowly or broadly defined; whether a conceptual framework was described; whether process and context evaluations were conducted; and finally the types of impact assessed (see Table 1).

Process evaluation is defined as evaluations which examine the extent to which the intervention was implemented as intended, including the distribution and coverage of its input components, such as availability of medicines, training of health workers, quality of care, as well as the acceptability of the intervention to the parties involved (Hawe *et al.* 2004; Oakley *et al.* 2006). It therefore helps to determine the internal validity of the evaluation, i.e. whether the intervention was adequately implemented and therefore the observed effects can be attributed to the intervention. In case of failure of an intervention, it helps to explain if the failure is due to an inherent problem with the intervention itself, i.e. the theory behind how it should work, or insufficient or inadequate ‘dose’ of implementation (Rychetnik *et al.* 2002; Schellenberg *et al.* 2004). Simply describing the intervention and the implementation process was therefore not considered a process evaluation, but rather information on the process (Rychetnik *et al.* 2002). This differentiation between process evaluation and information on the process was captured by two separate variables (see Table 1).

Context evaluation is defined as systematic documentation of naturally occurring events in the settings where the intervention was evaluated that might influence either positively or negatively the uptake of the intervention or the level of its impact. They are normally conducted throughout the evaluation period, or before and after, and are usually collected through

Table 1 Description of the variables describing the nature of the evaluation

Variable	Description
Nature of the evaluations	
1 Evaluation with a broad focus? (yes or no)	If the evaluation objective was not explicitly restricted, e.g. in the title or abstract, to a confined question, e.g. effect of the intervention on equity, or waiting time.
2 Conceptual framework? (yes or no)	If the study includes a section or diagram or reference describing how the intervention is theorized to work and how it will lead to the expected impact.
3 Information on the process? (yes or no)	If a brief description of the intervention and the implementation process was provided, without discussion on whether the intervention was adequately implemented, see process evaluation.
4 Process evaluation? (yes or no)	If factors determining a successful and adequate implementation process were conceptualized and measured, and discussion on the extent to which they explain or justify the hypothesized and observed intervention's effects was provided.
5 Information on context? (yes or no)	If at least some information on the context was included, whether or not a discussion on their implications was provided, see context evaluation.
6 Context evaluation? (yes or no)	If contextual variables are described, with information or reference to the tools used to collect contextual factors, and discussion on the extent to which evaluators believe they may have or not influenced the effectiveness of the intervention.
Nature of impact assessment	
1 Health outcomes only? (yes or no)	If the evaluation was limited to assessing the impact of the intervention on individuals' or populations' health (e.g. morbidity or mortality), and did not look at the impact on any of the health system (HS) building blocks.
2 HS effects on one targeted building block? (yes or no)	If the evaluation explored the impact on one HS building block targeted by the intervention. Health outcomes may or may not have been assessed.
3 HS effects on two targeted building blocks? (yes or no)	If the evaluation explored the impact on two HS building blocks targeted by the intervention. Health outcomes may or may not have been assessed.
4 HS impact across three or more building blocks, whether or not targeted by the intervention? (yes or no)	If the evaluation explored the possible impact on three or more HS building blocks, whether or not they were targeted by the intervention. Health outcomes may or may not have been assessed.
5 HS effects on other building blocks not targeted by the intervention? (yes or no)	This variable only looks at whether the evaluation explored the impact of the intervention on other HS building blocks not targeted by the intervention, regardless of which other effects from the above were also assessed.
6 HS effects on other sectors? (yes or no)	This variable only looks at whether the evaluation explored the impact of the intervention outside the health sector, regardless of which other effects from the above were also assessed.
7 Complex adaptive systems characteristics? (yes or no)	This variable looks at whether the evaluation attempted to capture effects linked to any of the characteristics of complex adaptive systems such as non-linearity of intervention effects or interaction between HS building blocks.

key informant interviews, or logs of relevant events, or interventions likely to affect the impact of the intervention in question (Hawe *et al.* 2004; Schellenberg *et al.* 2004). If some information on the context was provided but not in a systematic pre-conceptualized manner, this information was captured separately (see Table 1).

For those evaluations considering the intervention's impact across three or more building blocks, a deeper assessment of the nature of these evaluations was conducted, including choice of study design, methodological approaches and what impact was assessed with what measures. It also included whether they took into account any of the characteristics of complex adaptive systems (CAS) in their research design or methods, such as non-linearity of effects, time delays or feedback between the different health system components (de Savigny and Adam 2009; Paina and Peters 2011). This was done by reading the methods section of these evaluations and screening them for any mentioning of CAS or approaches to account for them, as described in de Savigny and Adam (2009) and Paina and Peters (2011).

Data coding was done separately by the two raters and any discrepancies were discussed until consensus was achieved.

A database of abstracted data was developed in Excel. Cross tabulation and frequencies were performed as well as an in-depth assessment of the nature of evaluations considering systems wide-effects, as described above.

Results

Study selection

The search in Medline and Embase resulted in a total of 2212 unique articles after removal of duplicates between the two databases, which accounted for 13%. Almost 60% of the articles were retrieved from Embase. Titles and abstracts were screened against the inclusion criteria and 91 articles were kept for data abstraction. Full text could not be located for 6 of those articles, leaving 85 articles for further analyses. The grey literature search resulted in 21 articles that met our inclusion criteria, retrieved from 7 out of 36 institutional websites (Table 2).

The majority of exclusions concerned studies that did not evaluate the output, outcome or impact of an intervention, but were situational analyses or cross-sectional surveys. Also

excluded were evaluations whose research objective was not concerned with the intervention's impact on the health system but on a clinical (was the treatment effective), technical (was a costing or monitoring tool effective) or operational aspect (was the training material applicable).

Eighty per cent of the evaluations were in low-income or lower-middle-income countries, with almost half of the evaluations conducted in sub-Saharan Africa (48%) followed by East Asia and the Pacific (19%) and Latin America and the Caribbean (11%) (see Figure 1).

Nature of the interventions

Out of the 106 evaluations included in this analysis, 91 were system-level interventions, targeting one or more of the health system building blocks; the remaining 15 were evaluations of the large scale-up of disease-specific interventions. Table 3 shows the types of interventions assessed by these evaluations and the most frequent examples within each type. Interventions centred around 11 major groups, with financing interventions being the most frequent, followed by models of service delivery, human resource strategies and scaling up of a health programme. HIV/AIDS was the most common disease

explored followed by malaria. In general, even when interventions were classified as system-level, the entry point was often a disease rather than strengthening of a particular aspect of the system across various health services. It is worth noting that interventions with the same name and overall objective varied substantially in the way they were defined (by the studies) with respect to their degree of complexity. For example, task shifting, voucher schemes and pay-for-performance involved 1–3, 3–5 and 1–3 building blocks, respectively (data not shown). The majority of interventions targeted the supply side; only a few focused on the demand side, e.g. using voucher schemes.

Table 4 shows the health system building blocks targeted by system-level interventions. So for example, 60 studies addressed service delivery, of which 20 focused on service delivery and one other building block. Of these 20 studies, 16 examined access, availability, timeliness, responsiveness or satisfaction; three evaluated public/private partnerships in service provision and four examined quality and safety of care. Most interventions were complex, targeting two or more building blocks. All building blocks were involved to a varying extent, although the most common intervention components were around service delivery, financing, health workforce and governance issues around service delivery. Information systems were the building block least targeted by system-level interventions.

Table 2 Sources of evaluations

Source	Organization/Database	Number included
Peer review literature	Medline and Embase	85
Grey literature	Global Development Network	11
	Global HIV/AIDS Initiative Network	2
	MEASURE Evaluation	2
	United States Agency for International Development (USAID)	2
	World Bank	2
	Massachusetts Institute of Technology (MIT) Poverty Action Lab	1
	University Research Corporation	1
	Total	21
Grand total		106

Nature of the evaluations

With respect to the nature of the evaluations, 43% have chosen broadly defined questions allowing for an assessment of a wider range of the intervention's effects (Table 5). Only half of the evaluations presented or referred to a conceptual framework, often linked to multivariate regression analyses of the intervention's impact on specific outcomes. The other half either listed a small set of questions or hypotheses that they aimed to answer or went directly to describe their data sources and findings without a prior description of which outcomes they chose to explore and why.

Around 60% of the evaluations provided information on the process of implementing the intervention and 20% provided contextual information to be able to situate the intervention and the observed effects within the context in which it was

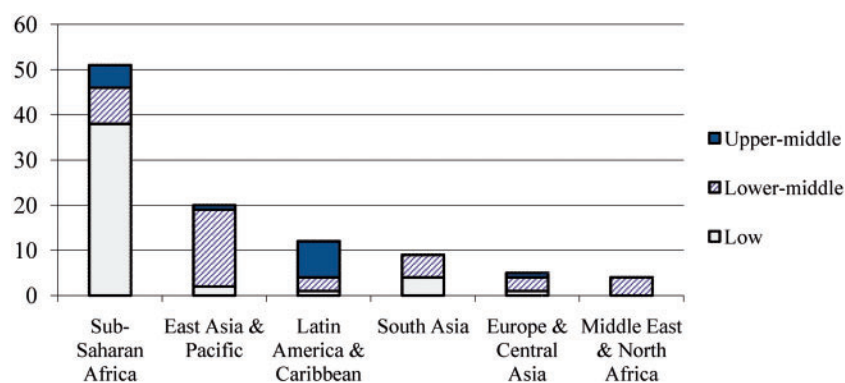


Figure 1 Location of evaluations^a

Note: ^aBased on World Bank 2011 country classification (World Bank 2011).

Table 3 Nature of the interventions

Types of interventions	Main examples	Scaling up, disease-specific (N)	System-level interventions, by number of building blocks involved						Total
			1	2	3	4	5	6	
Financing interventions (resource generation, pooling and payment mechanisms)	Health insurance (13), CCT (4), pay for performance (6), vouchers (3), user fees (2), others (10)	1	7	12	13	3	2	0	38
Models of service delivery	Community-based delivery (5), integration of care (7), others (4)	2	0	3	7	3	1	0	16
Human resource strategies	Task shifting (7), others (5)	0	2	7	3	0	0	0	12
Scaling-up	HIV/AIDS (10), malaria (1)	11	0	0	0	0	0	0	11
Public-private partnerships	Public-private mix for TB DOTS (2), contracting with NGOs (2), others (5)	0	1	4	3	0	1	0	9
Quality improvement	For PMTCT (2), others (5)	0	0	4	0	1	1	1	7
Sector reforms	Abolition of user fees (2), others (3)	0	0	0	3	2	0	0	5
Disease control	Malaria (2), NTDs (1)	1	1	0	1	0	0	0	3
Decentralization	HIV/AIDS (1), leprosy (1) services	0	0	1	0	1	0	0	2
Information systems	HMIS for HIV/AIDS (2)	0	1	1	0	0	0	0	2
Provider training	Twinning programme (1)	0	0	0	1	0	0	0	1
Total		15	12	32	31	10	5	1	106

Notes: CCT: conditional cash transfer; HMIS: health management information system; NGO: non-governmental organization; NTDs: neglected tropical diseases; PMTCT: prevention of maternal to child transmission of HIV/AIDS; TB DOTS: tuberculosis directly observed treatment short-course.

being implemented. With respect to process and context evaluations, 24% and 9% have included or referred to these components in their evaluations, respectively. Most of these evaluations assessed the scaling of HIV/AIDS services or the impact of global health initiatives on health systems and most were obtained from grey literature.

Despite the high proportion of studies that involved complex interventions (i.e. interventions which addressed multiple building blocks), the nature of the evaluations and the type of impact assessed did not reflect that complexity. Six evaluations looked only at health outcomes, e.g. mortality rates or treatment outcomes, while they evaluated interventions addressing up to five health system building blocks. More than half assessed the intervention's effects on one building block only, while the interventions involved were mostly complex with components covering two or more building blocks. Only seven explored the intervention's impact across three or more building blocks (Table 5).

Of the 19 evaluations that explored the intervention's impact on other building blocks, all except one looked at one other building block, most often service delivery. Only one study looked at the impact outside the health sector, in the form of household behaviour related to child labour and schooling, and employment of adults. It also looked at health outcomes but did not look at any of the building blocks targeted by the intervention (Rocha and Soares 2010). Finally, none of the evaluations explored system effects that reflect the complex adaptive nature of health systems.

In-depth assessment of evaluations that explored impact on three or more health system building blocks

We now turn our attention to the seven evaluations that explored the intervention's impact across multiple building blocks to explore the full range of system-wide effects they considered. Table 6 describes their main characteristics and methodological approaches. Six evaluations used mixed methods and one used only quantitative methods. In most cases, a limited set of commonly used effects measures was used. Plausibility designs with historical controls were the most frequent design choices (Victora *et al.* 2004). Among these evaluations, Loevinsohn *et al.* (2009) provide a good example of evaluations asking broader and multi-faceted questions, i.e. assessing the adequacy of the intervention on utilization, satisfaction of communities, quality of care, cost and efficiency. Similarly, Celletti *et al.* (2010) explored a comprehensive set of effects covering regulatory frameworks and reorganization of clinical services related to task shifting of HIV services to address health workforce shortage (Celletti *et al.* 2010). However, in most cases, the range of effects explored was limited, perhaps linked to the fact that conceptual frameworks were not always elaborate or comprehensive and often only limited to hypothesis testing (Loevinsohn *et al.* 2009; Witter *et al.* 2010). Interestingly the two most comprehensive evaluations, in our view, involved evaluations of interventions using participatory approaches in designing, monitoring and continuously improving the intervention, using data-driven and

Table 4 Number of building blocks targeted by system-level interventions and their sub-components^a

id	Health system building blocks and their sub-components	Number of building blocks targeted by the intervention						Total
		1	2	3	4	5	6	
1.0	Service delivery	0	20	25	9	5	1	60
1.1	Access, availability, timeliness, responsiveness, satisfaction	0	16	21	8	4	0	50
1.2	Public-private partnerships around service provision	0	3	9	1	2	1	15
1.3	Quality and safety of care provided	0	4	4	5	1	1	15
2.0	Health workforce	2	12	16	7	5	1	43
2.1	Supply and distribution	2	9	7	4	0	1	23
2.2	Personnel management and performance systems	1	0	3	3	1	0	8
2.3	Training (pre-service and in-service)	2	8	12	5	4	1	32
3	Information systems	1	2	1	3	3	1	11
3.1	Health information systems	1	0	0	2	1	1	5
3.2	Management information systems	0	2	1	2	2	0	7
4	Medical products, vaccines and technologies	0	1	12	6	4	1	24
4.1	Access and rational use	0	0	11	5	2	1	19
4.2	Public-private partnerships around medicines and technologies	0	1	3	0	1	0	5
4.3	Supply management	0	0	3	2	1	0	6
4.4	Quality and safety of medicines	0	0	0	0	1	1	2
5.0	Financing	7	13	18	5	4	1	48
5.1	Revenue collection and pooling	6	5	9	3	1	0	24
5.2	Payment mechanisms: provider	1	7	4	2	3	0	17
5.3	Payment mechanisms: beneficiary	0	1	6	1	1	0	9
5.4	Resource allocation	0	3	4	1	2	1	11
6	Governance	2	16	21	10	4	1	54
6.1	Level of decision making, institutional arrangements, accountability	0	6	7	5	1	0	19
6.2	Scope and location of service providers	1	8	15	5	3	1	33
6.3	Consumer and stakeholder involvement	1	2	3	3	2	0	11
	Total^b	12	32	31	10	5	1	

Notes: ^aAdapted from the description of the WHO health system building blocks (World Health Organization 2007; World Health Organization 2000) and the McMaster Health Forum, health systems evidence taxonomy (<http://www.healthsystemsevidence.org>).

^bThese do not include the 15 evaluations of large scale up of disease-specific intervention as the primary focus of the intervention is not a health system or sub-system building block.

participatory mechanisms (Doherty *et al.* 2009; Youngleson *et al.* 2010).

Journals publishing peer-reviewed studies

Finally, we also analysed the nature of the peer-review journals that published the evaluations included in our study. Overall, journals that accepted evaluations looking at a wider range of impact have also accepted those with a narrowly defined focus. However, evaluations that did not explore, or explored a limited set of, the intervention's impact on the health system, were mostly published in medical or specialized journals, while most evaluations that explored impact across multiple building blocks were published in journals focusing on health policy, public health or social sciences.

Discussion

In this paper, we reviewed recent evaluations of health systems strengthening interventions in LMICs to assess whether they

explored the intervention's effects across multiple health system building blocks, and if they did, to what extent, and with what methodological approaches.

Most of the evaluated interventions were complex, with 75% of them involving two or more health system building blocks. However, less than half of the evaluations asked a broad set of research questions to allow for a wider assessment of the intervention's impact on the health system. Only half presented or referred to a conceptual framework to guide the assessment of the intervention's impact. Less than a quarter included process evaluation and 9% included context evaluations.

Among those who conducted process evaluation, most have used classic indicators of the intervention's coverage or implementation rates, e.g. number or percentage of health workers trained, education sessions held, medicines kits distributed, etc. As Howe *et al.* (2002) argued, while it is logical to measure if what was promised actually happened, a more prudent approach is to also think through and examine the intervention's causal assumptions that may have led to the measured degree of implementation and impact, which may or may not

Table 5 Nature of evaluations compared with the level of complexity of the interventions (according to number of building blocks targeted)

Variables ^a	Scaling up, disease-specific	System level: Number of building blocks targeted by the intervention						Total
		1	2	3	4	5	6	
Number of evaluations	15	12	32	31	10	5	1	106
Number of evaluations with the following study objectives and methodological approaches								
Broad objective	9	3	14	11	6	2	1	46 (43%)
Conceptual framework	3	6	20	16	6	1	1	53 (50%)
Information on process	9	8	14	19	8	4	1	63 (59%)
Process evaluation	3	2	8	9	2	0	1	25 (24%)
Information on context	2	2	5	7	4	0	1	21 (20%)
Context evaluation	1	1	3	4	0	0	1	10 (9%)
Number of evaluations according to nature of impact assessment^b								
Health outcomes only	0	1	3	1	0	1	0	6 (6%)
HS impact on one targeted building block	3	9	20	17	3	3	0	55 (52%)
HS impact on two targeted building blocks	12	n.a.	6	11	6	1	0	36 (34%)
HS impact on three or more building blocks, whether or not targeted by the intervention	0	0	3	2	1		1	7 (7%)
Number of evaluations exploring the following effects								
HS impact on other building blocks not targeted by the intervention	0	4	12	2	1	0	0	19 (18%)
Impact on non-health sector	0	0	1	0	0	0	0	1 (1%)
CAS characteristics	0	0	0	0	0	0	0	0 (0%)

Notes: CAS: complex adaptive system; HS: health system; n.a.: not applicable.

^aSee Table 1 for a definition of the variables.

^bThe remaining two studies explored the intervention's impact on one other building block but not that targeted by the intervention. They are included in the variable on health system impact on other building blocks.

match the hypothesized theories that guided the intervention's design (Hawe *et al.* 2004).

Assessing the degree to which context evaluation has been adequately performed was much harder to undertake. This may be partly a reflection of the lack of guidance on how to take the impact of context into account and on how to report it, which made several studies stop at listing what else is happening, without attempting to evaluate their likely impact on affecting the course of the intervention and its applicability to other contexts (Rychetnik *et al.* 2002). Some studies only mentioned contextual factors in the discussion section to explain ex-post why the intervention did not work as intended or why the results were not as expected (Arifeen *et al.* 2009).

With respect to health systems impact, half of the evaluations assessed the intervention's impact on one targeted building block. Only seven explored the impact on three or more building blocks. One evaluation assessed the intervention's impact on other sectors. None explored the relationship and interconnectedness between the different building blocks or other characteristics of complex systems such as non-linearity of effects or time delays (Shiell *et al.* 2008).

Interestingly the two most comprehensive evaluations, in our view, involved evaluations of interventions using participatory approaches in designing, monitoring and continuously improving the intervention (Doherty *et al.* 2009; Youngleson *et al.* 2010). This may be something inherent to participatory

evaluations that led to a more comprehensive 'system-wide' approach to assessing the intervention's impact. For example, involving stakeholders early on in the design process and engaging them in assessing and solving implementation barriers is at the heart of systems thinking, where the intervention's effects, anticipated or not, can be explored, discussed and considered in designing and evaluating health interventions (de Savigny and Adam 2009).

Our findings are consistent with other similar studies. For example, Paina and Peters (2011) did not identify any examples where models of scaling up health interventions have been examined through the lens of complex adaptive systems (Paina and Peters 2011). Our findings are also consistent with our previous analysis, which could only identify few examples of comprehensive evaluations that considered the complexity and dynamic nature of health systems, all of which targeted specific diseases or conditions, e.g. tobacco control or obesity (de Savigny and Adam 2009).

We do not argue that all evaluations should be comprehensive. Indeed, in our report on systems thinking and its role in evaluations, we argued that not all interventions require a systems thinking approach. However, we argued that interventions can be seen as a continuum, where the more complex the interventions are, the more the need for systems thinking and comprehensive assessment of system-wide effects (de Savigny and Adam 2009).

Table 6 Main features of evaluations of system-wide effects, $N = 7$

Intervention description	Study design	Counterfactual	Building blocks targeted by intervention	Building blocks evaluated	Tools and analytical approaches	Effects measured
Participatory quality improvement programme for PMTCT in South Africa (Doherty <i>et al.</i> 2009)	Plausibility	Historical control	2	3	<i>Quantitative:</i> descriptive statistics <i>Qualitative:</i> structured interviews, direct observation	Quality of care; availability of medical staff; availability of key resources; supervision; access to services; utilization of services; community satisfaction
Contracting-in with a local NGO to manage the basic health units in one district, Punjab, Pakistan (Loevinsohn <i>et al.</i> 2009)	Probability (RCT)	Intervention/ comparison, and historical control	2	4	<i>Quantitative:</i> logistic regression of HMIS data and household surveys <i>Qualitative:</i> health facility surveys, interviews, observations of patient-provider interactions	Quality of care; availability of medical staff; availability of medicines; facility infrastructure; utilization of services; patient satisfaction; cost of care
Task shifting of HIV services involving community health workers: a multi-country study (Cellesti <i>et al.</i> 2010)	Comparative cross-country analysis	None	2	3	<i>Quantitative:</i> descriptive statistics <i>Qualitative:</i> key informant interviews, focus group discussions, in-depth interviews and review of key documents	<i>Regulatory framework:</i> collaborative planning, education and training, recruitment procedures, employment conditions, sustainability, informal constitution of CHWs. <i>Reorganization of clinical services:</i> competencies and scope of practices, access to services and quality, mentoring and supervision, service users' satisfaction, CHW views and opinions.
Community-based integrated delivery health interventions in seven research sites in Cameroon, Nigeria and Uganda (Remme 2010)	Probability (RCT)	Intervention/ comparison, and historical control	3	4	<i>Quantitative:</i> Chi square test for coverage indicators; cost analysis <i>Qualitative:</i> in-depth interviews, focus group discussions, key informant interviews, structured observation	Utilization of services; provider costs; community involvement, staff shortages
Abolition of user fees for primary care in Zambia (Masiye <i>et al.</i> 2010)	Plausibility	Historical control	3	4	<i>Quantitative:</i> descriptive statistics <i>Qualitative:</i> patient survey and provider interviews	Utilization of services; patient perception of quality of care (availability of drugs, staff courtesy, waiting time, facility cleanliness, overall responsiveness); provider perception of service provision (staff workload and morale, quality of care, quality of relationships between facility administrators and district health management teams)
Exemption policy for free delivery and caesarean in Senegal (Witter <i>et al.</i> 2010)	Plausibility	Historical control	4	4	<i>Quantitative:</i> descriptive statistics, costing and cost-effectiveness analysis <i>Qualitative:</i> key informant interviews, focus group discussions, in-depth interviews and review of key documents	Utilization rates; quality of care; health workforce motivation; costs and cost-effectiveness; total expenditure; supply management; equity
Improving PMTCT through quality improvement intervention in South Africa (Youngleson <i>et al.</i> 2010)	Plausibility	Historical control	6	3	<i>Quantitative:</i> time series analysis	Utilization of services; use of medicines; quality of health information data

Notes: CHW: community health worker; HMIS: health management information system; PMTCT: prevention of maternal to child transmission of HIV/AIDS; RCT: randomized controlled trial.

This study highlights the need to understand the possible barriers to more comprehensive evaluations, when they are appropriate. A recent study eliciting the views of a wide range of stakeholders in the Eastern Mediterranean Region identified a range of barriers to more comprehensive evaluations. They included lack of technical capacity to undertake such evaluations; limited awareness and appreciation of the value of adopting a more comprehensive, systems thinking approach, in designing and evaluating health systems interventions; as well as a perceived notion of their costliness, combined with limited support from, and investments by, research funders. Respondents also highlighted the importance of generating awareness among policy makers to provide the necessary support and demand for such comprehensive evaluations (El-Jardali *et al.* forthcoming).

Our analysis has a number of limitations. First, it only includes evaluations available from two literature databases, Medline and Embase, and a limited number of web-based grey literature. Second, the analysis only focused on evaluations published in 2009–10. However, our aim was not to take stock of all evaluations undertaken on this topic, rather to have a general understanding of how the field of evaluations has been developing, particularly in response to recent calls for more rigorous and comprehensive assessment of efforts to strengthen health systems, including the application of systems thinking concepts and tools in conceptualizing and evaluating health interventions.

Conclusion

Very few evaluations attempted to conceptualize the possible effects of interventions on multiple health system building blocks. While we do not argue that all interventions require a comprehensive evaluation of the system-wide impact, we argue for the need for more evaluations that explore the wider range of impact on the health system as a whole, and even beyond the health sector, as appropriate.

There are several untapped resources that could make significant contribution to this field, including consideration of the underlying concepts of complex adaptive systems; systems thinking concepts, tools and approaches; as well as adopting and learning from social sciences and policy analysis perspectives, both involving complex social and political phenomena constructed and influenced by human action, all very relevant to health systems and the field of evaluations (de Savigny and Adam 2009; Gilson *et al.* 2011; Paina and Peters 2011).

Finally, this study highlights the need to strike a balance between identifying easy-to-answer research questions, vs asking more difficult but important research questions. The latter would require adopting a more problem-solving attitude to research and being more flexible and innovative in employing research strategies that are deemed appropriate for the research questions (Paina and Peters 2011).

Acknowledgements

The authors wish to thank Dominique Guinot, Barbara Koloshuk and Kasper Wyss for their contribution to initial phases of this study.

Funding

This work was supported by the Alliance for Health Policy and Systems Research, World Health Organization. The named authors alone are responsible for the views expressed in this publication and they do not necessarily represent the views, decisions or policies of the World Health Organization.

Conflict of interest

None declared.

Supplementary Data

Supplementary data are available at *Health Policy and Planning* Online.

References

- Adam T, Ahmad S, Bigdeli M, Ghaffar A, Røttingen J-A. 2011. Trends in health policy and systems research over the past decade: still too little capacity in low-income countries. *Plos One* **6**: e27263.
- Alliance for Health Policy and Systems Research. 2011. *Implementation Research Platform*. Online at: <http://www.who.int/alliance-hpsr/projects/implementationresearch/en/index.html>, accessed 30 August 2012.
- Arifeen SE, Hoque DME, Akter T *et al.* 2009. Effect of the Integrated Management of Childhood Illness strategy on childhood mortality and nutrition in a rural area in Bangladesh: a cluster randomised trial. *The Lancet* **374**: 393–403.
- Bennett S, Adam T, Zarowsky C *et al.* 2008. From Mexico to Mali: progress in health policy and systems research. *The Lancet* **372**: 1571–8.
- Bennett S, Agyepong IA, Sheikh K *et al.* 2011. Building the field of health policy and systems research: an agenda for action. *Plos Medicine* **8**(8): e1001081.
- Best A, Clark PI, Leischow SJ, Trochim WM. 2007. *Greater Than the Sum: Systems Thinking in Tobacco Control*. Bethesda, MD: National Cancer Institute, US Department of Health and Human Services, National Institutes of Health.
- Butland B, Jebb S, Kopelman P *et al.* 2007. *Foresight: Tackling Obesity: Future Choices*. London: UK Government Office for Science.
- Celletti F, Wright A, Palen J *et al.* 2010. Can the deployment of community health workers for the delivery of HIV services represent an effective and sustainable response to health workforce shortages? Results of a multicountry study. *AIDS* **24**: S45–57.
- Cohen J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* **20**: 37–46.
- de Savigny D, Adam T. 2009. *Systems Thinking for Health Systems Strengthening*. Geneva: Alliance for Health Policy and Systems Research, World Health Organization.
- Doherty T, Chopra M, Nsiband D, Mngoma D. 2009. Improving the coverage of the PMTCT programme through a participatory quality improvement intervention in South Africa. *BMC Public Health* **9**: 406.
- El-Jardali F, Adam T, Ataya N, Jamal D, Jaafar M. Constraints to applying systems thinking concepts in health systems: a regional perspective from Eastern Mediterranean countries. Forthcoming.
- Evans T, Nishtar S, Atun R, Etienne C. 2008. Scaling up research and learning for health systems: time to act. *The Lancet* **372**: 1529–31.
- Gilson L, Hanson K, Sheikh K *et al.* 2011. Building the field of health policy and systems research: social science matters. *Plos Medicine* **8**(8): e1001079.

- Hawe P, Shiell A, Riley T, Gold L. 2004. Methods for exploring implementation variation and local context within a cluster randomised community intervention trial. *Journal of Epidemiology & Community Health* **58**: 788–93.
- Landis JR, Koch GG. 1977. Measurement of observer agreement for categorical data. *Biometrics* **33**: 159–74.
- Leischow SJ, Best A, Trochim WM *et al.* 2008. Systems thinking to improve the public's health. *American Journal of Preventive Medicine* **35**(2 Suppl.):S196–203.
- Leischow SJ, Milstein B. 2006. Systems thinking and modeling for public health practice. *American Journal of Public Health* **96**: 403–5.
- Lewin S, Lavis JN, Oxman AD *et al.* 2008. Alma-Ata: Rebirth and revision 2 – Supporting the delivery of cost-effective interventions in primary health-care systems in low-income and middle-income countries: an overview of systematic reviews. *The Lancet* **372**: 928–39.
- Loevinsohn B, ul Haq I, Couffinhal A, Pande A. 2009. Contracting-in management to strengthen publicly financed primary health services—The experience of Punjab, Pakistan. *Health Policy* **91**: 17–23.
- Masiye F, Chitah BM, McIntyre D. 2010. From targeted exemptions to user fee abolition in health care: Experience from rural Zambia. *Social Science & Medicine* **71**: 743–50.
- Mills A. 2012. Health policy and systems research: defining the terrain; identifying the methods. *Health Policy and Planning* **27**: 1–7.
- Mills A, Gilson L, Hanson K, Palmer N, Lagarde M. 2008. What do we mean by rigorous health-systems research? *The Lancet* **372**: 1527–9.
- Oakley A, Strange V, Bonell C, Allen E, Stephenson J. 2006. Process evaluation in randomised controlled trials of complex interventions. *British Medical Journal* **332**: 413–6.
- Paina L, Peters DH. 2011. Understanding pathways for scaling up health services through the lens of complex adaptive systems. *Health Policy and Planning* **27**: 365–73.
- Remme JHF. 2010. Community-directed interventions for priority health problems in Africa: results of a multicountry study. *Bulletin of the World Health Organization* **88**: 509–18.
- Remme JHF, Adam T, Becerra-Posada F *et al.* 2010. Defining research to improve health systems. *Plos Medicine* **7**(11): e1001000.
- Ritz LS, Adam T, Laing R. 2010. A bibliometric study of publication patterns in access to medicines research in developing countries. *Southern Medical Review* **3**: 2–6.
- Rocha R, Soares RR. 2010. Evaluating the impact of community-based health interventions: evidence from Brazil's family health program. *Health Economics* **19**: 126–58.
- Rychetnik L, Frommer M, Hawe P, Shiell A. 2002. Criteria for evaluating evidence on public health interventions. *Journal of Epidemiology and Community Health* **56**: 119–27.
- Savedoff WD, Levine R, Birdsall N. 2006. *When Will We Ever Learn? Improving lives through impact evaluation*. Washington, DC: Centre for Global Development.
- Schellenberg JRM, Adam T, Mshinda H *et al.* 2004. Effectiveness and cost of facility-based Integrated Management of Childhood Illness (IMCI) in Tanzania. *The Lancet* **364**: 1583–94.
- Shiell A, Hawe P, Gold L. 2008. Complex interventions or complex systems? Implications for health economic evaluation. *British Medical Journal* **336**: 1281–3.
- Swanson RC, Bongiovanni A, Bradley E *et al.* 2010. Toward a consensus on guiding principles for health systems strengthening. *Plos Medicine* **7**(12): e1000385.
- Swanson RC, Mosley H, Sanders D *et al.* 2009. Call for global health-systems impact assessments. *The Lancet* **374**: 433–5.
- Trochim WM, Cabrera DA, Milstein B, Gallagher RS, Leischow SJ. 2006. Practical challenges of systems thinking and modeling in public health. *American Journal of Public Health* **96**: 538–46.
- van Etten G, Baltussen R, Bijlmakers L, Niessen L. 2006. Advancing the Mexico agenda for health systems research – from clinical efficacy to population health. *Tropical Medicine & International Health* **11**: 1145–6.
- Victora CG, Habicht JP, Bryce J. 2004. Evidence-based public health: moving beyond randomized trials. *American Journal of Public Health* **94**: 400–5.
- Witter S, Dieng T, Mbengue D, Moreira I, De Brouwere V. 2010. The national free delivery and caesarean policy in Senegal: evaluating process and outcomes. *Health Policy and Planning* **25**: 384–92.
- World Bank. 2011. *World Development Indicators 2011*. Washington, DC: World Bank.
- World Health Organization. 2000. *The World Health Report 2000: Health systems: Improving performance*. Geneva: World Health Organization.
- World Health Organization. 2005. *World Health Assembly Resolution (WHA58.34) Ministerial Summit on Health Research*. Geneva: World Health Organization.
- World Health Organization. 2007. *Everybody's Business: Strengthening Health Systems to Improve Health Outcomes: WHO's Framework for Action*. Geneva: World health Organization.
- Youngleson MS, Nkurunziza P, Jennings K *et al.* 2010. Improving a mother to child HIV transmission programme through health system redesign: quality improvement, protocol adjustment and resource addition. *Plos One* **5**(11): e13891.