Professional accountability for women's, children's and adolescents' health: what mechanisms and processes are used, what works?

A systematic literature review

The Partnership for Maternal, Newborn & Child Health

Table of contents

Acknowlegements

This systematic review was commissioned by the Partnership for Maternal, Newborn & Child Health (PMNCH) as a PMNCH White Paper by Andrea Nove, Adriane Martin Hilber and Carolyn Blake (Novametrics, Inc). Miriam Lewis Sabin, Manager, Accountability (PMNCH) coordinated the preparation of this white paper, with the assistance of Yajna Moloo (PMNCH Consultant), under the leadership of Helga Fogstad, Executive Director (PMNCH).

Cover photo credit: Jessica Lea/DFID.

Copy editing, design and layout by Genève Design.

Abbreviations and acronyms

EmONC	emergency obstetric and newborn care
EWEC	Every Woman Every Child
GRADE	Grading of Recommendations, Assessment, Development and Evaluations
HICs	high-income countries
LMICs	low- and middle-income countries
MDR	maternal death review
MNH	maternal and newborn health
NGO	non-governmental organization
PAHO	Pan American Health Organization / WHO Regional Office for the Americas
PBF	performance-based financing
PMNCH	Partnership for Maternal, Newborn and Child Health
PPH	postpartum haemorrhage
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QI	quality improvement
RBF	results-based financing
RCT	randomized controlled trials
SDG	sustainable development goal
UAF	Unified Accountability Framework
UN	United Nations
USAID	United States Agency for International Development
WCAH	women's, children's and adolescents' health
WHO	World Health Organization

Executive summary

Background

Achieving the Sustainable Development Goals (SDG) for women's children's and adolescents' health (WCAH) requires strengthening accountability processes within health systems to increase performance and quality of services, and ultimately improve population health outcomes. Guidance, however, on how best to implement accountability mechanisms for WCAH, is lacking. To better inform WCAH policies and programmes, this paper presents the scope and characteristics of, and evidence behind professional accountability mechanisms aimed at improving health system performance. To ensure learning across a variety of contexts, this global review includes evidence from high-, middle- and low-income settings.

Methods

This systematic review was based on findings from a preliminary scoping of the field of accountability in women's, children's and adolescents' health, and the results of a previous structured review, which focused on accountability for maternal and newborn health in sub-Saharan Africa. *(1)* The search terms for this systematic review expanded the search terms from the previous review to include studies from low- and middle-income countries (LMICs) outside of Africa and studies on accountability mechanisms for child health, adolescent health, reproductive health and nutrition. The initial screening resulted in 4390 papers, published between 2008 and 2018. Of these, 36 peer-reviewed papers were considered to be within the scope of this review. Papers were categorized according to six types of mechanisms that were assessed as being under-reviewed in the literature: 1) accreditation/certification/standard setting; 2) benchmarking/scorecards; 3) clinical audits; 4) clinical/supportive supervision; 5) managerial audit/supervision; and 6) results based financing (RBF). A framework was developed to facilitate the analysis of the strength and sustainability of the accountability mechanisms under review, as well as their evaluation.

Results

The majority of the papers included in this review focused on women's health (mainly maternal health and noncommunicable diseases), with only a minority looking at children's and/or newborn health, and only one focusing on adolescent health. Overall, there was a fairly even split between studies conducted at a national, sub-national and local level. Hardly any of the papers described a recourse mechanism (such as remedial actions, remedies or sanctions for non-performance) when performance was found to be substandard, and few described equity effects.

A large majority of studies evaluated the role of accountability mechanisms for improvement in quality of care. Few considered morbidity and mortality as an outcome of interest, and instead focused on process outcomes. Factors for success in general were identified as follows, but differ according to the type of mechanism: involvement of a broad range of stakeholders, a culture of learning, the voluntary nature of participation in the mechanism, integration into national processes, champions and local ownership within the system, independent and external assessors, data transparency, clear recommendations, and annual reviews and dialogue.

Conclusion

This review assesses the strengths and limitations of professional accountability mechanisms and facilitates the disentanglement of complex "accountability ecosystems" through a health systems performance and governance lens. The review indicates that a multidisciplinary approach to accountability in practice is essential to sustain improvements. Likewise, there is a need for greater understanding of and clarity as to how accountability mechanisms (rather than specific tools or approaches) operate within a health systems improvement cycle, who should be engaged to lead and participate in such a process, and how success should be measured and ultimately sustained.

Background

Sustainable Development Goal (SDG) 16 emphasizes the need "to build effective accountability and inclusive institutions" to achieve all of the SDGs and this is particularly true for those related to the health sector. *(2)* Accountability by governments for their global and national commitments to women's, children's and adolescents' health are increasingly monitored, measured and scrutinized at all levels as governments strive to meet the challenge of accelerating progress under the SDGs. *(3,4)* The UN Secretary General's Every Woman Every Child (EWEC) *Global Strategy for Women's, Children's and Adolescents' Health* defines how governments can deliver on their commitments and is underpinned by a Unified Accountability Framework (UAF) that aims to strengthen effective accountability processes and mechanisms at country, regional and global levels across sexual, reproductive, maternal, newborn, child and adolescent health interventions. *(3,5)*

Accountability can be increased via "bottom-up" or "short-route" approaches whereby users become aware of their rights, build expectation and become effective observers and monitors of the providers and the services they request. "Long-route" or "top-down" accountability is also needed to strengthen the governance of the health system by making managers and policymakers more answerable for the quantity, quality and efficiency of the system, for sustainable change. There has been considerable investment in strengthening health system governance, and on accountability to advance global and national goals. However, there is limited evidence of accountability mechanisms within evidence-based interventions or their contribution to the implementation of health interventions. There are notable exceptions related to improving access to information to monitor, for example, health finances, performance or political commitments; or supply-side interventions such as human resource and financial management systems through accreditation. *(6–8)*

Global guidance on evidence-based interventions has rarely included specific processes to call to account those charged with implementing and supervising the interventions or assuring specific outcomes. Monitoring and evaluation measures have largely focused on determining the relative success of the interventions on higher level outcomes, rather than on whether the system has been strengthened by the integration of an accountability mechanism to ensure the improvements are sustained over time – particularly in contexts where the use of data for decision-making is limited. Understanding is lacking as to how, when and in which contexts specific accountability mechanisms can be put in place within health system interventions, despite the broad field of study on health system policy and planning and governance processes. Answerability in a public health system can be equated with the responsiveness of those in charge within a health system (for example, providers, managers or district officials) to act when standards, norms, and quality of services or health outcomes are inadequate or neglected. While governments are supposed to provide oversight, supervision (answerability) and if necessary, sanctions when service norms and standards are not met, in reality such mechanisms for accountability are weak, poorly implemented, or non-existent in many countries. *(9)*

Heightened concern by citizens and communities about gaps in service standards and quality in the health sector, aided by greater awareness of health rights, has increased scrutiny and calls for greater accountability of the health system and providers specifically. There is growing consensus that while all approaches can advance some degree of accountability, the entire "accountability ecosystem" – the context, actors and their interactions – will play a role in how effective the accountability mechanism will ultimately be. The degree to which various accountability mechanisms (e.g. public, professional, financial, legal) and approaches reinforce each other will likely determine their sustainability and success.

Although there are several ways of categorizing types of accountability, this paper uses the categorization developed by Brinkerhoff that includes: professional (an aspect of performance) accountability, financial accountability, and social and political accountability. *(10,11)* These categories align with the "three Rs" used by the UAF and Partnership for Maternal Newborn and Child Health (PMNCH): Results, Resources and Rights. Within each category of accountability, different types of accountability mechanisms may operate, and many are most successful when they are employed in tandem. Some types of accountability (such as social accountability) have been systematically studied. *(12–14)* However, less is known about the effectiveness of other types of accountability mechanisms, especially in the sphere of WCAH. One area that has not been systematically studied is professional accountability as a key aspect of health system performance for WCAH, which encompasses performance measurement and evaluation, service delivery improvement and management reform.

In this paper, we use the following commonly accepted definitions of accountability and accountability mechanisms:

Accountability exists when "… an individual or body, and the performance of tasks or functions by that individual or body, are subject to another's oversight, direction or request that they provide information or justification for their actions." *(15)* In other words, the individual or body is answerable to another individual or body for the results of their action or inaction, and there are consequences – positive or negative – for decisions that are made.

An **accountability mechanism/process** is a formalized process that aims to ensure accountability. In relation to professional/performance accountability, examples include: accreditation/certification, benchmarking, clinical audit, management audit and performance-based financing. An accountability tool (for example a scorecard or checklist) may be part of an accountability mechanism/process but is not in itself a mechanism/process.

This review set out to answer the following questions:

1. What types of accountability mechanisms and processes are used to improve accountability for the professionalism/performance of health workers or health systems, in relation to women's and/or children's and/or adolescents' health?
2. Which of these have been shown to lead to improvements in mortality, morbidity and/or quality of WCAH care?
3. Which have been shown not to contribute to improved outcomes, or have been shown to have had unintended consequences?
4. What are the characteristics of an effective professional/performance accountability mechanism/process in the sphere of WCAH? Which accountability mechanisms/processes have been shown to be sustainable, durable, generalizable and/or scalable? That is, what is the potential for their institutionalisation? What is the potential for learning across different contexts?

Methods

Initial scoping exercise

An initial scoping exercise was done to take stock of the literature available on professional/performance accountability in advance of developing the search criteria for the systematic review. The exercise aimed to:

a) assess whether there has been any previous systematic or other type of review of accountability processes, mechanisms and tools as they relate to women's and/or children's and/or adolescents' health (WCAH) – including any aspects of health such as nutrition, immunization, noncommunicable diseases, mental health, sexual and reproductive health (SRH), maternal and newborn health (MNH), health education, and/or communicable diseases – that provided detailed evidence or lack thereof of 'what works/doesn't work' in the implementation of accountability mechanisms;

b) inform the decision about the search terms and databases to be used for the systematic review to be conducted; and

c) identify other aspects of accountability which have been under-investigated but which have sufficient information to warrant review.

We included all reviews (grey and peer reviewed) that describe accountability mechanisms, processes or tools that have been implemented and evaluated, measured or assessed in some way.

It was important for the design of the review to complement rather than repeat existing work in this area. This involved reviewing theoretical and conceptual literature as well as assessments and evaluations of aspects or reasons behind successful interventions.

Although the objective of the scoping exercise was to focus the systematic review itself on 'what works/doesn't work', it was important to ensure that the search terms used were sufficiently broad to pick up all the relevant types of accountability that had been previously reviewed. An initial collection of key reference documents and recent PMNCH resources were provided by WHO colleagues in the Mother, Child and Adolescent Health department and the Reproductive Health and Research department. Following this initial collection of suggested articles and reviews, we conducted a further search in Google Scholar and PubMed to locate other reviews and conceptual articles that met the search criteria for the scoping exercise. Key search terms included:

a. accountability mechanisms processes tools systematic review

b. budget tracking systematic review

c. political accountability systematic review

d. legal accountability systematic review

e. democratic accountability systematic review

f. professional accountability systematic review

g. accountability for quality of health care systematic reviews.

Based on this scoping exercise, a protocol for the systematic review was developed.

Search strategy for the systematic review

As a point of departure, the same search strategy was used as a previous structured review, which focused on accountability related to maternal and newborn health in sub-Saharan Africa. *(1)* The search terms were then expanded to include studies from LMICs outside of Africa and studies relating to accountability mechanisms for child health, adolescent health, reproductive health and nutrition.

In an extraction grid, we recorded information including: type of mechanism, tools used, existence of recourse mechanism and any information that might be relevant to the development of a set of accountability markers and assessment criteria (e.g. enabling factors and barriers to sustainable implementation of the mechanism). This review indicated that the adapted Brinkerhoff typology of mechanisms used in the 2016 IJGO) paper was still applicable.

Using the Brinkerhoff classification, we were able to classify the articles found. Most of the reviewed papers related to performance accountability mechanisms, but all types of mechanisms except political accountability were represented in the review. The vast majority were about maternal and newborn health (MNH), with a few about reproductive, child or adolescent health. Most of the studies took place in sub-Saharan African countries, with a few from South Asia and a few global studies.

In addition to this preliminary search, we also searched for reviews of the various accountability mechanisms noted earlier. The search for existing literature reviews on professional/performance accountability in health found that there are six main types of mechanisms: 1) accreditation/certification of health facilities, 2) benchmarking, 3) clinical audits (including death reviews and audits), 4) clinical/supportive supervision, 5) managerial audit/supervision and 6) results-based financing (RBF) or performance-based financing (PBF). These are defined as follows:

Accreditation usually involves a voluntary programme, sponsored by a nongovernmental organization (NGO), in which trained external peer reviewers evaluate a healthcare organization's compliance and compare it with pre-established performance standards. *(16)*

Benchmarking is a process of comparative evaluation and identification of the underlying causes leading to high levels of performance. It involves a sustained effort to measure outcomes, compare these outcomes against those of other organizations to learn how those outcomes were achieved, and to apply the lessons learned in order to improve. *(17,18)*

Clinical audit involves measuring care against specific criteria, taking action to improve it if necessary, and monitoring the process to sustain improvement. *(19)*

Clinical supervision involves monitoring the activities of health workers; whereas **supportive supervision** is a non-authoritarian process of helping staff to improve their own performance by encouraging open, two-way communication and using a team approach to facilitate problem solving. *(20)*

Managerial audit/supervision involves senior colleagues monitoring the activities of more junior health workers and recommending improvements where needed. This often involves rural or isolated health workers receiving visits from managers based at district offices. *(21)*

Results-based financing is defined as "a cash payment or non-monetary transfer made to a national or sub-national government, manager, provider, payer or consumer of services after predefined results have been attained and verified". *(16)*

Performance-based financing is one of the most common supply-side incentive schemes being tested in low- and middle-income countries to improve health systems through systemic change, and is often focused on quality of care improvements. *(22)* If fully implemented, PBF involves changing institutional roles, responsiveness and enforcement mechanisms to motivate providers to change professional behaviour positively to improve health system and organizational performance. *(23)*

All six types of mechanisms were considered for inclusion in this systematic review. The scoping exercise located existing relevant reviews for managerial audit and RBF/PBF, and the key findings from these reviews are summarized below. For the remaining four types of performance accountability mechanisms, we carried out a systematic review of the literature, which is described later.

Thematic accountability areas not included in the systematic review

MANAGERIAL AUDIT: FINDINGS OF PREVIOUS REVIEWS

A 2012 Cochrane review of the effect of managerial audit and feedback on healthcare outcomes *(26)* concluded that this mechanism is associated with "small but potentially important" improvements in professionals' adherence to practice standards (e.g. proper use of treatments or laboratory tests, or improving the overall management of patients with chronic disease), and that it may be most effective when baseline performance is low, the auditor is a supervisor or colleague (as opposed to an external auditor), the process occurs more than once and/or in both written and verbal formats, and when the outputs of the process include explicit targets and a plan of action. However, the quality of the evidence for the effect of the mechanism on patient outcomes (as opposed to compliance with clinical standards) was judged to be low. Based on earlier research, the authors hypothesized that the reason for the effect being generally small is that, whilst audit and feedback can be effective in motivating health workers to improve standards, the lack of an enabling environment may limit their capacity to do so.

RESULTS-BASED AND PERFORMANCE-BASED FINANCING: FINDINGS OF PREVIOUS REVIEWS

This summary covers articles that reported on supply-side PBF initiatives aimed at improving health worker performance, as that was the focus of this study. It does not include the literature on demand-side initiatives, which aimed to bring about other types of change. Given the substantial body of evidence on RBF, this review summarizes only reviews (peer-reviewed and one significant grey literature review).

We located seven recent (2011–2017) systematic literature reviews and summaries of the evidence related to RBF. *(27–33)* All focused on low- and middle-income countries (LMICs) and often reviewed multiple outcomes, usually including quality of service improvements and cost. Four of the articles were included in the synthesis below as they summarized key components of the PBF literature such as how PBF is monitored, what aspects of quality it can affect, and evidence gaps, among others. (22, 34–36) Characteristics of the PBF reviews synthesis are described in Table 1. The synthesis described subsequently has informed the review but has not been formally included in this systematic literature review.

Table 1. Performance-based financing review characteristics

Thematic focus	Outcomes reviewed				
	Total (n=7)	Quality of care	Cost-effective-ness	Access (equity)	Service utilisations
Maternal, newborn, child health *(27,28)*	2	2	2	1	1
HIV/AIDS *(29)*	1	1	1	1	
Primary health care *(30)*	1	1			1
General health *(31–33)*	3	3	2	2	1
Year of publication		**Review quality**			
2011	1	High — Cochrane review			
2013	2	Low — Grey literature review		Moderate — Systematic review	
2014	1	Moderate — Systematic review			
2016	2	Moderate — Systematic review			
2017	1	Moderate — Systematic review			

These reviews indicate that while RBF is appealing, results can be context specific. It is attractive because it can be targeted to address quality of care shortfalls by linking performance with strategic incentives, and as such achieves accountability with a carrot (reward) rather than a stick (sanction). *(34)* RBF can be targeted to specific bottlenecks and priorities, which is a major reason it is considered to have great potential. It has been applied most frequently to improve MCH within primary health care services, and can be applied in very different settings including those in post-conflict. RBF is often led by external players (e.g. the World Bank, the Global Fund or USAID) though a few countries such as Rwanda have included it as part of their national health policy funded by the national budget. *(27)*

In recent years, there has been a plethora of articles and reviews of PBF as a performance accountability mechanism to improve quality of care, including a Cochrane review in 2011 *(30)* and an evidence gap mapping study currently being undertaken. Indeed, this review likely did not capture all relevant articles on this topic. *(35)* Recent reviews have cited some effectiveness of PBF in improving use and quality of maternal and child health services and access and use of HIV-related services. *(28,29)* For example, one study found PBF associated with improved quality of care for patients living with HIV by reducing rates of attrition and treatment failure, yet these findings were not replicated in other studies. *(29)* Specifically, improvements have been documented on process quality indicators in antenatal care including: adherence to protocols, availability of skilled staff (Burundi, Democratic Republic of Congo), drugs and provider knowledge (Burundi, Haiti, Egypt), utilization and coverage of maternal services (India, Kenya, Uganda), institutional deliveries (Rwanda), and uptake of modern family planning (Burundi). *(27,28)*

In some countries however, provision of financial incentives for performance improvements in quality of care was insufficient to change behaviour of providers and may result in negative outcomes on structural quality, such as a decrease in the level of availability of equipment and drugs; neglect of untargeted services; or provision of needless or detrimental services and fraud. *(27,28)* It was also unclear what the long-term effect of PBF would be on provider behaviours and expectations, possibly crowding out inherent staff motivation. *(27)* PBF targeting HIV service delivery improved quality and access to services, but concern was raised that these achievements may have been at the cost of other health priorities through negative spill over effects of other services. *(29)* Evidence of improvements for people in the poorest socioeconomic quintiles was mixed compared to those in the relatively better off groups, [32] with some reviewers suggesting that a combination of different approaches used in combination could have promising effects on equity and quality of services, but this remains under-studied. *(27)*

More information is needed to assess the impact of PBF on WCAH and its causal pathways in LMICs. (27,28,31) Evidence is weak on value for money of PBF interventions, and particularly the connections between costs and effects or outcomes. *(30,33)* Study designs were also weak, *(33)* leading many reviewers to conclude that despite promising potential, evidence on PBF is insufficient to date. *(37)* In addition, many studies were conducted by those implementing PBF, which could raise the issue of conflict of interest and bias for the results obtained. *(33)*

Strategies for success cited in the PBF literature focused on the strength and quality of the verification of the performance improvement measures. While the evidence is largely inconclusive thus far, it can be seen that PBF can improve quality and utilisation of services in the short term but evidence of systemic and structural changes needed to sustain such improvements remains scarce.

It should be noted that a Cochrane review of the effect of maternal, perinatal and child death reviews and audits on mortality rates is under way but not yet published. *(24)* A systematic review of the use of criterion-based clinical audit to improve the quality of obstetric care was published in 2010 *(25)*, but this was somewhat narrower in scope than our research questions so this type of mechanism was included in our review. Similarly, a review of clinical supervision was published in 2011 *(21)*, but this focused on primary health care in low- and middle-income countries only. While it was rather narrower in scope than our research questions, it was nevertheless included in the review.

Data sources

We searched PubMed, CINAHL, Web of Science, Cochrane Library and Virtual Regional Portal for PAHO. All but the last of these were selected because they featured most heavily in the literature identified via the initial scoping exercise. The Virtual Regional Portal for PAHO was selected to ensure appropriate representation of literature written in Spanish and Portuguese. In addition, we scanned the reference lists of papers identified via the above searches to identify other relevant studies. The literature search was limited to studies involving human subjects.

Search strategy

The initial scoping exercise helped to identify the words and phrases commonly used when referring to accountability generally, and performance/professional accountability specifically, and these informed the development of search terms, which can be found in the annex.

Study selection

Literature research results (citations and abstracts) were exported to reference management software to identify and remove duplicate records. They were also extracted to a spreadsheet to facilitate the title and abstract review. Each title and abstract was screened according to the following inclusion criteria:

1. Published in 2008 or later, and
2. Published in English, French, Spanish or Portuguese, and
3. The title or abstract described or mentioned one or more of the following performance accountability mechanisms or processes: accreditation/certification, benchmarking, clinical audit, clinical supervision, and
4. The title or abstract indicated that there was an assessment of the effect that the mechanism or process had on mortality, morbidity and/or quality of care for women, children (including newborns) and/or adolescents.

A team of two researchers were briefed by one of the review authors to read all titles and abstracts in the spreadsheet, and allocate each to one of the following categories:

- Meets inclusion criteria
- Query (no abstract, or insufficient information in the abstract to be sure of eligibility)
- Does not meet inclusion criteria and is not relevant
- Does not meet inclusion criteria but is a relevant literature review
- Does not meet inclusion criteria but is another type of relevant/interesting article
- Duplicate record not previously identified.

As a calibration exercise, the first 100 titles and abstracts were reviewed independently by both researchers and one of the review authors, then the three of them compared their results and discussed those for which their assessments did not match. They agreed on these before continuing with the title and abstract review.

Every title/abstract classed by the researchers as 'include' or 'query' was reviewed by two of the review authors to decide whether or not they met the inclusion criteria. If necessary, the full article was accessed for this purpose. Differences of opinion between the review authors were resolved by discussion, and reasons for exclusion were recorded. A random sample of 10% of the excluded titles/abstracts was reviewed by one of the review authors to check the classification. No systematic errors were identified as a result of this quality check.

Data extraction

The review authors extracted data in duplicate from each article that was judged to meet the inclusion criteria and populated a standardized extraction grid (Table 2).

Table 2. Data items in extraction grid

Year of publication
Year(s) of accountability intervention
Language
Researcher ID
Eligibility
Region(s) and/or country/ies, income group(s)
Study design / data collection method
Theoretical framework used (if any)
Exposure-outcome pairing*
Element of women's, children's and adolescents' health
Thematic focus
Risk of bias
Study quality *(38)*
Type of accountability process, system, or mechanism within performance accountability
Study setting (administrative level and health system level)
Actor type(s)
Rights-holder(s) and duty-bearers
Aim of the intervention / problem that it aimed to solve
Aims of study
Stage(s) of the accountability mechanism reached
Recourse mechanism: Existed? Used? Impactful?
Key findings by theme: what works / what does not work
Conclusions (What strategies led to which outcomes?)
Equity effects reported
Generalizability: To what extent are the study findings generalizable to other settings?
Implications for policy and practice

* *In this context, the 'exposure' was the accountability mechanism itself, and the 'outcome' was mortality, morbidity and/or quality of care. If the study assessed more than one outcome, new rows were added to the extraction grid to examine each pairing separately.*

As a calibration exercise prior to the full text review, three papers were each reviewed by two of the review authors, and the extraction grid populated in duplicate. All three review authors discussed the results and resolved differences of opinion via discussion. None of the reviewers was blind to the journal titles or to the study authors or institutions.

Analysis framework

A systematic narrative synthesis was conducted, exploring the relationship and findings both within and between the included studies. Numerical frequencies were calculated wherever applicable. A narrative synthesis was considered to be a more appropriate approach than a meta-analysis because: (a) the scope of the study was broad: it included several different types of mechanism that did not have common outcomes, and (b) much of the evidence (especially about enabling factors) was qualitative: given its objectives, it was important for the study to report on enabling factors as well as on quantitative effect sizes.

The Every Woman Every Child UAF considers three stages of an accountability mechanism: monitor, review and remedial action. This framework was used during the analysis of the identified papers, by considering which stage each mechanism reached during its implementation. We also sought information from the published papers about what happened before and after implementation, e.g. which stakeholders were engaged in the design and launch of the mechanism, did it result in sustained changes to policies, processes, practices or norms, (and therefore how sustainable was the change attributed to the mechanism)? An analysis framework was developed to assess the implementation pathway achieved through the accountability intervention from inception and planning (including stakeholder engagement and consultation), through monitoring, reviewing and acting on a specific accountability mechanism, to finally how the action was sustained within the health system. The analysis framework facilitated contextualizing how the accountability mechanism was or was not embedded in a broader process or accountability ecosystem necessary for sustainability. *(9)* The framework hypothesizes that a functional accountability mechanism will involve key stakeholders at the outset in the pre-implementation phase (e.g. to ensure that the power brokers are supportive from the start); and will work with the stakeholders through the implementation phase (i.e. monitor, review, act). It further hypothesizes that for an accountability mechanism to be effective, it must not only be implemented but also lead to change in health system norms and professional culture, and eventually a sustained increase in accountability of the professional system's performance through institutionalization and transformation. The distinction between institutionalization and transformation is that institutionalization occurs when there are sustained, country-driven changes to *processes and practices*, whereas transformation occurs when there are changes to *norms and/or policies.*

Strengths and limitations of the review

This review employed a robust methodology involving a wide range of data sources to capture as much relevant literature as possible from the last 10 years, including non-English language publications (eventually, only one of the reviewed papers was not in English, which was perhaps due in part to the relatively small number of search terms used in French, Spanish and Portuguese). It aimed to fill an important gap in the literature by focusing on professional accountability mechanisms, and specifically on those that have not recently been systematically reviewed. However, the review was limited by the exclusion of the wide range of grey literature published on the topic in recent years.

The strength of the conclusions that can be drawn from a systematic review depend heavily on the strength of the evidence provided by the studies included in the review. Most of the studies included in this report are rated as low quality because they are largely observational in design, and many do not include control or comparison groups, which limits the extent to which the observed change can be attributed to the mechanism. In this report, we draw some tentative conclusions about what may work in performance and professional accountability based on these studies, but it should be noted that these conclusions are based on weak evidence, and that stronger evidence would be required before this report can be used to make global recommendations on how professional/performance accountability mechanisms can be employed to improve WCAH. However, this paper can serve as background to the preparation of normative guidance and global recommendations as part of a GRADE exercise: a method of assessing the certainty in evidence and the strength of recommendations in health care interventions.

This review included studies from high-, middle- and low-income countries, which is a strength, but does affect the interpretation of the results. As noted below, the literature on clinical audits and clinical/supportive supervision is mostly from low- and middle-income countries, while the literature on other types of mechanisms is mostly from high-income countries in Europe. This limits the extent to which we can draw conclusions about the applicability of different mechanisms to different settings and the extent to which we can assume that 'what works' in one setting will work in another.

The focus of this review on WCAH proved to be both a strength and a limitation. The achievement of the Global Strategy for WCAH requires a specific focus on these population groups, so the focus on them in this review is appropriate to ensure that issues that are specific to these groups are not masked by issues affecting the general population. Furthermore, this review has highlighted the paucity of studies relating to accountability of those providing health services to newborns, children and adolescents, who have specific health needs and often lack a voice within discussions about their needs. On the other hand, it is important to consider WCAH within the wider health system to avoid siloed thinking, so these results need to be interpreted within this wider context.

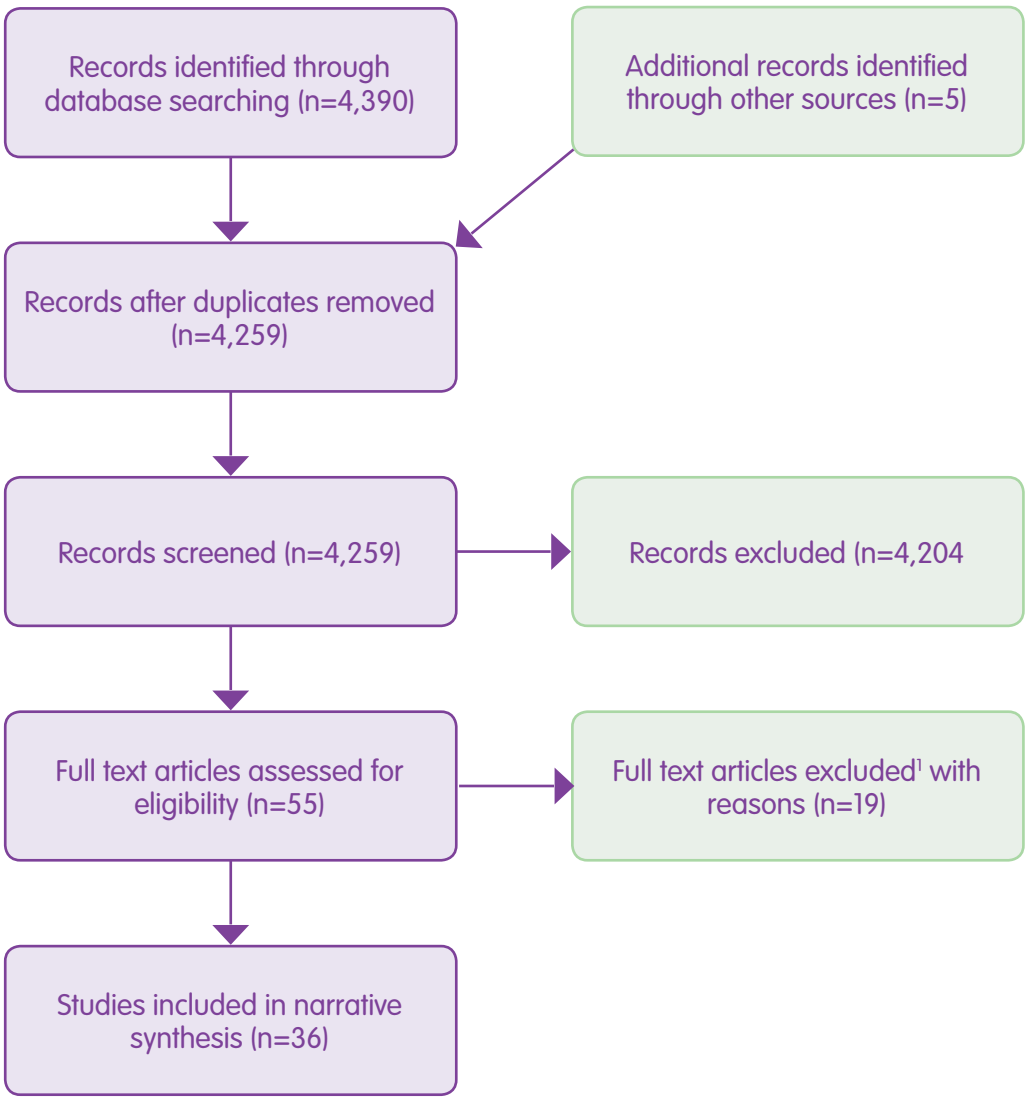
Similarly, limiting the review to studies that used quality of care, mortality or morbidity as an outcome measure led to the exclusion of many potentially relevant studies, e.g. those looking at other important outcomes or simply describing promising approaches to accountability.

The attempt to include six different types of mechanism within the same study has the advantage of comprehensiveness, but limited the amount of detail provided about each type of mechanism in this report. It may be appropriate to look in more detail at one or more of the mechanism types as a next step in this process.

Results

Study selection

Fig. 1. PRISMA flow diagram



1 Reasons for exclusion available upon request

Study characteristics

The PRISMA flow diagram in Fig. 1 depicts the study selection process of the systematic review, which resulted in the inclusion of 36 studies in the narrative synthesis. Of the 36 papers about professional/performance accountability mechanisms for WCAH, 27 described clinical audit mechanisms, 11 described accreditation/certification/standard-setting mechanisms, 7 described benchmarking/scorecard mechanisms, and 4 described clinical/supportive supervision mechanisms. (This adds up to more than 36 because some of the papers were about interventions that included more than one of these types of mechanism.)

Table 3 shows the characteristics of the 36 studies included in this systematic review. Despite efforts to include papers in languages other than English, only one non-English-language paper met the inclusion criteria. Almost half of the papers were published in or after 2014, indicating an increase in the number of papers published on this topic in recent years. Due to the large number of observational studies included in the review, study quality was generally rated as 'low' according to the GRADE criteria. *(38)* Just 7 studies included a control or comparison sample in the study design, so that observed changes at the intervention sites could be compared against non-intervention sites.

About half of the studies were conducted in low- and middle-income settings (mostly African countries), but the vast majority of these were about clinical audits or clinical/supportive supervision mechanisms. Most of the studies on other types of performance accountability mechanism were carried out in high-income countries, mostly in Europe.

The studies tended to take a quantitative approach and relied on retrospective analysis of data. The majority of the papers focused on women's health (mainly maternal health and noncommunicable diseases), with a few looking at children's and/or newborn health, and only one focusing on adolescent health. Overall, there was a fairly even split between studies conducted at a national, sub-national and local level. However, for obvious reasons, there were no reports of benchmarking mechanisms operating at a local level, nor of clinical supervision mechanisms operating at a national level. Nearly all of the mechanisms operated at secondary or tertiary levels of the health system rather than at lower levels. There was a tendency to focus on the public sector: of the 21 papers indicating the sector in which the mechanism operated, 15 were public sector, 3 were private sector and 3 covered both public and private sectors.

A variety of actors were described for each of the mechanisms. As expected, nearly every paper mentioned the involvement of facility-based health workers, and half of the papers mentioned health facility managers. Professional associations were involved in most of the accreditation and benchmarking mechanisms, but were less often mentioned in the papers about clinical audit and clinical/supportive supervision. Conversely, health ministries were fairly often a key actor for clinical audit mechanisms, but rarely for accreditation or benchmarking mechanisms. It is notable that patients/service users were identified as a key actor in only 9 of the 36 studies, most of which were about accreditation mechanisms. It is, of course, possible that service users were actors in more of the mechanisms, but that this fact was not mentioned in the published paper.

Most of the studied professional accountability mechanisms achieved implementation (or the "action" stage of the EWEC Unified Accountability Framework, but a quarter of them did not. Hardly any of the papers described a recourse mechanism (e.g. remedial action, remedy, sanction for non-performance) when performance was found to be substandard, and few described equity effects such as the impact of the mechanism on population sub-groups or levels of the health system.

There is increasing interest in professional accountability shown by the increase in studies since 2008 with 16 of the reviewed studies having been conducted since 2014. PBF reviews have also increased with 4 of the 7 reviewed having been undertaken in the same period.

Table 3. Study characteristics

	Total (n=36)	Type of mechanism			
		Accreditation/ certification/ standard-setting (n=11)	Bench-marking / scorecard (n=7)	Clinical audit (n=27)	Clinical/ supportive supervision (n=4)
Type of mechanism					
Accreditation/ certification/ standard setting	11	11	4	5	0
Benchmarking/scorecard	7	4	7	4	1
Clinical audit	27	5	4	27	1
Clinical/supportive supervision	4	0	1	1	4
Language					
English	35	11	7	26	4
French	1	0	0	1	0
Year of publication					
2008-2010	9	1	1	6	2
2011-2013	11	4	3	9	0
2014-2017	16	6	3	12	2
Study quality					
High	1	0	0	0	1
Moderate	8	1	1	8	0
Low or very low	27	10	6	19	3
Setting					
High-income	17	10	6	11	1
Low- and middle-income	20	1	2	17	3
Africa	16	0	1	14	3
Americas	5	3	1	2	1
Europe	12	6	5	9	0
South-East Asia	4	1	1	3	0
Western Pacific	1	1	0	1	0
Study design					
Quantitative	29	10	7	22	3
Qualitative	1	0	0	1	0
Mixed methods	6	1	0	4	1
Prospective	11	4	3	7	1
Retrospective	25	7	4	20	3
Control or comparison sample?	7	2	0	5	1
Element of WCAH					
Women's health	28	7	4	24	2
Newborn health	5	1	3	3	2
Children's health	5	3	2	1	1
Adolescent health	1	1	1	0	0
Thematic focus					
Maternal newborn health	24	2	2	20	3
Noncommunicable diseases	9	6	5	6	1
Secondary health care	4	1	1	3	0
Sexual and reproductive health and rights	2	1	0	1	1
General health	2	2	0	0	0
Primary health care	1	0	1	0	1

	Total (n=36)	Type of mechanism			
		Accreditation/ certification/ standard-setting (n=11)	Bench-marking / scorecard (n=7)	Clinical audit (n=27)	Clinical/ supportive supervision (n=4)
Administrative level					
Multi-country	2	1	2	2	0
National	11	4	2	8	0
Sub-national/district	12	3	3	9	2
Community/local	11	3	0	8	2
Health system level					
Secondary/tertiary	34	10	7	25	3
Primary	7	1	1	4	1
Community/outreach	1	0	0	1	0
Not clear	1	1	0	1	0
Sector					
Public	15	4	3	11	2
Private	3	2	0	1	0
Both	3	0	0	3	0
Not clear	15	5	4	12	2
Actors					
Health workers: facility	35	10	7	27	4
Health facility managers	17	9	3	11	1
Professional association	10	6	5	7	0
Health ministry	9	0	1	7	2
Patients/service users	9	7	2	5	1
Programme implementers	7	0	0	7	0
NGO	7	2	2	4	2
Champions	4	1	0	3	1
International donors	4	1	0	3	0
Local government	3	1	1	1	1
District Health Management Team or similar	3	0	1	2	1
Health council/hospital board	2	0	2	2	0
Traditional leaders	1	0	0	1	0
Health workers: community	1	0	0	1	0
Stage of UAF reached					
Monitor	1	1	0	0	0
Review	8	3	2	6	0
Act	20	1	3	16	4
Institutionalization	6	6	2	4	0
Transformation	0	0	0	0	0
Not clear	1	0	0	0	0
Recourse mechanism					
Yes	4	1	0	3	1
No	1	0	0	1	0
Not clear	31	10	7	23	3
Equity effects reported					
Any	6	2	0	5	0
None	30	9	7	22	4

Note: *Some studies had more than one of the listed characteristics, which is why some of the column totals in the above table add up to more than the total number of papers.*

The impact of professional/performance accountability mechanisms on mortality, morbidity and quality of care in women's, children's and adolescents' health, and strategies for success

Accreditation/certification and standard setting of health facilities

Of the 11 papers discussing accreditation/certification and standard setting mechanisms:

- 4 papers focused solely on accreditation and standard setting *(39–42)*
- 7 papers discussed this mechanism in combination with benchmarking/scorecards, *(43–46)* or audit-related mechanisms. *(47–49)*

Each paper described a single exposure-outcome pairing. All papers focused on assessing the effect of the mechanism on quality of care. None presented morbidity or mortality outcome data.

Of the 11 papers, five demonstrated that accreditation schemes could improve compliance with diagnostic and treatment protocols for breast cancer care, *(39,45,46,49)* and assisted reproductive technologies. *(48)* Strategies for success included: strong reporting and accountability requirements, favouring disease-specific accreditation schemes over a more general scheme (here breast cancer over general cancer focus), *(39)* transparent and clear standards, regular inspections and a consistency in the approach. *(48)* The voluntary nature of the accreditation programme was seen as an important success factor and a means of learning in a non-blame environment. *(45,46)* Peer audits and reviews by external experts on an annual basis were also important factors. *(45,46)* In one study, hiring a dedicated staff member (in this instance a medical oncology nurse practitioner) was acknowledged as an important intervention component. *(49)* In another, strict screening and vetting of organizations before being able to join the system, and strict criteria for receiving this accreditation were seen as important, as well as the continuous adaptation of benchmarking criteria based on practice, and changes in the evidence behind guidelines. *(46)*

Two papers showed that standard setting/accreditation initiatives could lead to improvements in health facility safety culture and patient safety in paediatric wards: one demonstrated that in a two-year timeframe, the scheme led to an estimated 68% decrease in the number of serious safety events, a reduction in medication errors with serious harm, and improved compliance with clinical protocols, *(40)* and the other found that although there was a significant decrease in the rate of prescription errors, administration errors did not change. *(42)* The timeframe however, was only a year, and these findings do not align with a different study, *(21)* which found a decrease in medication administration errors after the same accreditation process. For Peterson et al, *(40)* strategies for success included: safety-based training for all staff, training in root cause analysis, failure mode classification of events and safety behaviour, integration of and collaboration between risk management and clinical staff, consistent coding and classification of serious safety events and adoption of multiple safety metrics, creating a new safety leadership infrastructure, and fostering transparency of data and safety event details. For Mekory et al, *(42)* successful strategies were pre-established explicit criteria or standards assessed by external evaluators, interviews with staff and patients, on-site observations of patient care processes, and training of staff to refresh knowledge of protocols as well as to enhance awareness to possible errors (pre-accreditation).

Two papers described quality improvement (QI) cycles that aimed to improve quality of care and update clinical guidelines. The first discusses QI in the context of neonatal intensive care units, and demonstrates that the initiative led to a broad set of changes including: the revision of guidelines for perinatal care at national level, the replacement of hand disinfectant in one unit, and the revision of oxygen saturation levels across participating institutions. *(44)* The second, however, concludes that the QI process with external peer review did lead to changes in treatment patterns and multi-disciplinary breast cancer care. *(47)* Strategies for success included: oversight of the process by a national neonatal network, monitoring of both process and outcome results on a continued basis, bi-annual meetings to discuss results across institutions and discuss problematic areas, and focus on a formative process in a non-blame environment.

One paper suggests that accreditation of health facilities can improve the availability of services. *(41)* The accreditation of private health facilities delivering abortion related services was associated with improved availability of essential equipment and drugs, opening hours, and transparent display of the availability of free services. Strategies for success included: public private partnerships, working through regional government and with NGOs, as well as subsidies to pay facilities for the costs of these services.

Benchmarking/scorecards

The six papers analysed in this section include:

- 3 papers about benchmarking within a broader accreditation/certification scheme *(44–46)*
- 2 papers about benchmarking within quality improvement schemes *(43,50)*
- 1 paper on benchmarking within criteria-based audits. *(51)*

The six papers described 7 exposure-outcome pairings (Table 4). All papers focused on quality of care, with one paper *(43)* also assessing the effect of the mechanism on morbidity.

Table 4. Seven exposure-outcome pairings examined in the six benchmarking /scorecard papers

Exposure (type of mechanism)	Outcome			Total
	Mortality	Morbidity	Quality of care	
Benchmarking within a broader accreditation/ certification scheme	0	0	3	3
Quality improvement schemes	0	1	2	3
Benchmarking within criteria-based audits	0	0	1	1
Total	0	1	6	7

The three papers that assessed a benchmarking initiative within an accreditation scheme on quality of care indicated that benchmarking could contribute to compliance with diagnostic and treatment protocols for breast cancer care *(45,46)* and neonatal intensive care. *(44)* Strategies for success included: oversight of the process by a national network, monitoring of both process and outcome results on a continuous basis, the voluntary nature of the system with a focus on a formative process in a non-blame environment, review meetings across institutions to discuss results and problematic areas, and benchmarking criteria being adapted on a continuous basis based on practice and changes in the evidence behind guidelines.

Of the two papers that looked into benchmarking within a QI scheme on quality of care, one assessed an intervention in the area of diabetes care, and demonstrated a significant improvement in quality screening and changes in management of cases. *(43)* This success was attributed to a non-punitive environment, dialogue between providers in quality meetings and anonymized scorecards when comparing data across hospitals. The other paper also shows improvements in health facility standards to provide emergency obstetric and newborn care (EmONC). *(50)* Success factors included the integration of the mechanism in national level government processes (although externally funded), the use of scorecards, district level multi-stakeholder meetings, and action plans developed and reviewed based on quarterly assessments.

Finally, one paper assessed a benchmarking component within a criteria-based audit intervention on quality of care, and found improvements in the diagnostic process and surgical treatment of women diagnosed with breast cancer. *(51)* The main success factor cited was the annual review of data that motivated health facilities to improve.

Clinical audits

The 27 papers analysed in this section include:

- 9 papers about criteria-based audit mechanisms, *(48,51–58)* and 3 about peer audit or review mechanisms *(46,47,59)*
- 9 papers about maternal or perinatal death review or audit mechanisms, *(60–68)* and 3 about confidential enquiries into maternal deaths *(69–71)*
- 3 papers about other types of mechanisms that involved an element of clinical audit (45,49,72)

The 27 papers described 36 exposure-outcome pairings (Table 5). The papers on criteria-based and peer audit mechanisms mainly attempted to assess the impact of the mechanism on quality of care, but a few also considered mortality and/or morbidity as outcomes. The papers on death reviews and confidential enquiries focused more on mortality as an outcome, but several also considered quality of care as a step on the causal pathway to reduced mortality. Just three papers used morbidity rates as an outcome.

Table 5. 36 exposure-outcome pairings examined in the 27 clinical audit papers

Exposure (type of mechanism)	Outcome			Total
	Mortality	Morbidity	Quality of care	
Criteria-based audit	2	1	9	12
Peer audit or review	0	1	3	4
Death review or audit	7*	0	6	13
Confidential enquiry	2	1	1	4
Other	1	0	2	3
Total	12	3	21	36

* *This includes one paper that considered both maternal and perinatal mortality as outcomes (61)*

Of the seven papers that measured changes in mortality rates after implementation of death reviews or audits, *(61,64–68)* five observed significant decreases in maternal and/or perinatal mortality after the mechanism was introduced. The exception was a 21-month intervention; *(66)* some of the other studies in this group noted that an improvement became apparent only after the mechanism had been operational for some time, which may in part explain this anomaly (although some of the other studies did observe a reduction in less than 21 months, so this cannot be the sole explanation). One of the five studies showing mortality reduction found that the reduction was evident only among women who had had a caesarean section: there was no significant difference between the intervention and control groups in terms of maternal mortality following vaginal birth. *(68)* Strategies for success included: regular supportive supervision visits to study sites to ensure compliance with the maternal death review (MDR) system, national support from opinion leaders, a no-blame culture, the process being led/owned by senior health workers (it is notable that one study found that recommendations made by the external research team were not implemented), *(66)* and a multidisciplinary approach. It was noted by several authors that facility-based death audits overlook deaths that occur in the community, and therefore give only a partial picture.

The two papers that examined the relationship between mortality and confidential enquiries into maternal deaths *(69,70)* both observed a decrease in institutional maternal mortality over the study period. One attributed this to the mechanism leading to improved clinical management of HIV which was the leading cause of maternal death in South Africa at the time. *(70)* Strategies for success included: independent assessors, clear recommendations, transparency, and national ownership of the mechanism.

Of the two papers about criteria-based audit that considered maternal mortality as an outcome, *(55,56)* one noted a reduced case fatality rate for women diagnosed with eclampsia. *(55)* Strategies contributing to this positive outcome were: the criteria were set by health workers in consideration of local context, and implementation of recommendations was the responsibility of a multidisciplinary team of health workers. The conclusions of the other paper were unclear.

The paper that examined the relationship between criteria-based audit and morbidity *(53)* concluded that implementation of this mechanism was associated with reduced incidence of severe postpartum haemorrhage (PPH) after vaginal delivery. In that study, the clinical guidelines focused on vaginal delivery, which is perhaps why no similar reduction was observed in the incidence of severe PPH after caesarean section. The paper that examined peer audit's relationship with morbidity *(59)* found no significant difference in incidence of severe PPH when comparing the intervention and control sites, but speculated that sample contamination may have confounded the comparison. Strategies for success included: a multidisciplinary approach, institutional support for the mechanism and a culture of learning rather than blame.

Six papers considered how maternal death reviews (MDRs)/audits may affect quality of care, *(60–65)* of which five concluded that MDR implementation was associated with improvements to clinical practice such as increased use of the partograph during labour. The exception was a study from Uganda, *(65)* which reported that although the MDR process yielded concrete recommendations about improvements to quality of care, resource constraints meant that they could not all be implemented during the study period. Strategies for success included: regular supportive supervision visits to study sites, support from national and/or local government, leadership of senior clinicians (champions), staff ownership of the process, a no-blame culture, and a multidisciplinary approach.

Nine papers looked at the relationship between criteria-based audit and quality of care, *(52–58)* mostly in relation to obstetric emergencies such as PPH and eclampsia. These studies tended to treat quality of care as a multidimensional concept and therefore measured it using a variety of criteria. Echoing the results from the earlier systematic review on this topic, *(25)* all found that at least some aspects of quality of care (as measured by adherence to established clinical guidelines) improved following the implementation of the mechanism, and several noted that improvements tended to be greatest for the criteria for which baseline performance was poor. Where improvements were not observed, this was attributed to poor communication to health workers about what was expected of them on that specific criterion *(52,57)* and/or to low staff motivation. *(57)* Strategies for success included: the criteria being set by the health workers in consideration of the local context, the time invested in orienting health workers to the importance of the mechanism, a multidisciplinary approach, health workers having a sense of ownership due to contribution to the design and implementation of the mechanism, institutional support, a culture of learning not blame, consistent application of the guidelines by an external auditing body, and regular supportive supervision visits to participating health facilities.

Of the three papers that examined the relationship between peer audit and quality of care, *(46,47,59)* two found no significant change in quality of care in relation to PPH and breast cancer. The other *(46)* found that minimum standards of care for breast cancer improved significantly for all dimensions that had room for improvement at baseline. Strategies for success included: requiring participating facilities to have good data systems, and using external rather than internal peers to carry out the audits.

Clinical/supportive supervision

Of the four papers that involved a clinical/supportive supervision mechanism, (50,56,73,74) all considered quality of care as an outcome, and one *(56)* also considered mortality (stillbirths), although the quality of this study was very low and the conclusions unclear. The remaining three papers all identified improvements in quality of care, defined as higher levels of adherence to protocols *(73,74)* and/or an increase in the number of health facilities achieving the standards required for EmONC accreditation. *(50)* One of the studies *(74)* assessed client satisfaction, and found that, although adherence to protocols improved, client satisfaction did not.

There is no consensus in these papers about what strategies may have contributed to the observed improvements in quality of care. One noted that the activities and support of a 'champion' (in this case the parent of a child patient) drove the process, *(73)* and another noted that having an on-site supervisor was helpful for the smooth running of the mechanism. *(74)* However, that same paper noted that a reliance on an on-site supervisor was insufficient for tackling the broader health system issues that were highlighted by the mechanism; a conclusion supported by the third paper, which concluded that the involvement of a broad range of stakeholders was an important ingredient for success. *(50)*

Discussion

Professional/performance accountability, as defined by Brinkerhoff as early as 2003, *(75) (76)* aims to raise the quality of service delivery care to higher legal, ethical, financial, and professional standards and is currently gaining traction. It implicates individual service providers to be answerable for the quality of their work and their adherence to standards and protocols; health services to oversee and enforce standards of practice from providers by putting in place accountability mechanisms that require providers and managers to take responsibility for their actions; and professional associations and other bodies to set standards and help to enforce them to the limit of their capacities. *(77)*

Increasing global calls for attention to professional/performance accountability from the SDGs related to WCAH and the expansion of universal health coverage is evident in the 36 articles reviewed, some of which describe multiple approaches. According to the literature, accountability for improving quality of WCAH care and reducing mortality and morbidity has so far focused largely on verifiable, criteria-based mechanisms such as clinical audit and accreditation/certification/standard-setting, which have been shown to be successful at improving quality of care in high-income countries (HICs) and have long been identified as critical to improving the quality of care in LMICs with high maternal and child mortality. *(78)* It has also focused almost exclusively on maternal and newborn health and not the broader WCAH agenda of the SDGs, still less other issues critical to this agenda such as nutrition. These mechanisms rely on top-down or "long route" approaches to accountability, which must be led by government or professional accrediting institutions where they exist. When successfully implemented, these approaches have a greater potential to achieve a higher level of accountability because they include remedial action in their design (higher level oversight and potential for sanction). Other professional/performance accountability mechanisms are more participatory in nature, mixing top-down and bottom-up ("short-route") approaches to accountability. The benchmarking/scorecard and clinical/supportive supervision approaches described in the literature focus on the health facility level, and on primary health care workers and their direct supervisors.

Despite the enthusiasm for such approaches, and their study, particularly in recent years (almost half of the papers were published in or after 2014), the evidence is largely drawn from observational studies, the majority without a comparison group, which do not command high quality ratings according to international GRADE criteria for systematic reviews. *(38)* Evidence for approaches not related to clinical audits or supervision is primarily drawn from HICs, where the health system context and oversight function is often more predictable and answerable to legal, ethical, financial and professional codes of conduct. The studies that were conducted in LMICs (mostly African countries) tend to describe clinical audits or clinical and supportive supervision mechanisms, where the accountability is hosted at a secondary or tertiary (sub-national) level that can be more readily facilitated and controlled than when higher national level engagement of the health sector is required. These accountability efforts can have an effect at the facility level, as demonstrated in a number of the studies.

The primary actors within the accountability mechanisms reported in the literature are professional associations (for accreditation and benchmarking) and the ministry of health for more clinical service-based interventions (e.g. audits, supervision). In less than a third of the studies, was the link to patients or users of the service evident. This finding is in line with the factors for success defined by the studies. Some of the key factors of success included: accountability processes that were supported by the system and internal champions, criteria-based verification and monitoring (that was internally established by those voluntarily participating in the process), and a non-judgemental and often confidential context. In a context where trust between providers and users of the health system is often tenuous, engaging with patients or users could be perceived as threatening. *(79)* Providers may feel that patients could potentially introduce blame and accusations, that could result in limiting transparency in sharing data and information needed for constructive

dialogue and recommendations that can be taken forward by those implicated in the accountability process. Changing the terms of engagement between patients and providers will likely require concerted efforts such as those being tried by public accountability stakeholders to bridge the perception and trust divide, before providers themselves and their supervisors truly seek partnership with patients around professional accountability. *(9)* Facility committees and other organizing structures to give citizens a voice in the health sector could facilitate partnership for accountability. But these also depend on many of the same success factors measured by providers: trust, transparency, confidentiality, voluntary participation, criteria based decision-making, and context-specific processes that have been agreed by all stakeholders. *(8)*

Professional/performance accountability mechanisms operate at the clinical and institutional level to effect change in the quality of care (among other desired outcomes). The reviewed articles provided some evidence that professional accountability mechanisms can contribute to improvements in quality of care, particularly adherence to standards that are on the change pathway for reductions in mortality and morbidity. Only death reviews and audit studies claimed to have an impact on mortality through remedial actions taken due to the accountability feedback loop. Equity effects were generally not studied, which is an important omission in the era of 'leave no one behind' as set out in the Global Strategy for WCAH and the SDGs. RBF showed some positive results in increasing access and use of services through user-based incentives and some provider incentives but the evidence was insufficient to know whether short term gains and improvements would be sustained once the incentives ended, and there was little indication that government could maintain the programmes without external funding. *(27)* Yet generally, outcome measures should be treated with some degree of caution as the studies were largely observational, often lacked randomization and controls, and sometimes demonstrated internal researcher bias. Attribution of the professional accountability interventions towards improvement in quality of care is equally challenging as few studies controlled for complementary inputs such as training, incentives and other factors that may have been working in parallel to improve the quality of care. Furthermore, improved health outcomes are not the only desirable outcome of an accountability mechanism: process indicators can also be important (and are more easily measured and potentially less susceptible to confounding).

As set out in Table 6, given the limitations mentioned, the degree of success in enhancing accountability of health worker and system performance through the various mechanisms reviewed in this paper seems to depend on: (a) the rules of the approach and their application in the process (transparency and appropriateness of the criteria and consistency upon which the mechanism is assessed); (b) who is leading the process (internal champions; yet external, non-aligned assessors); (c) whether staff feel it is voluntary and within a no-blame culture; (d) support by the health system through facilitating dialogue, engaging diverse stakeholders (i.e. multidisciplinary approaches); and (e) whether health workers feel supported for the implementation of recommendations. Accreditation processes in HICs seem also to create greater accountability when health workers are provided with additional training.

Table 6. Key ingredients for success in professional accountability processes

	Accreditation, certification, standard setting	Benchmarking, scorecards	Clinical audits	Clinical/ supportive supervision	Managerial audit	Performance based financing
Strong reporting and verification requirements (consistency and criteria-based oversight)	X	X		X		X
Supervision, monitoring and discussion of the mechanism on a regular basis	X	X	X	X		X
Involvement of a broad range of stakeholders; multidisciplinary		X	X	X		
Culture of learning/no blame	X	X	X			
Voluntary nature of the engagement with the mechanism	X	X				
National and institutional support; integration of mechanism into national processes (i.e. action plans, strategies etc.)	X	X	X			X
Champion (senior staff) to drive the process; process led internally by senior health workers/managers (local ownership of the process)	X	X	X	X	X	
Independent (external) assessors	X		X			X
Transparency of data, yet anonymized for improved dialogue		X	X		X	X
Clear criteria on protocols (established by health workers that know local context) and continuously updated	X	X	X			X
Clear recommendations and implementation led locally		X	X		X	
Annual review of data to motivate health facilities and workers	X	X			X	

This review uses the EWEC UAF and its antecedents (stakeholder engagement, design) and post effects (sustainability and transformative change) to assess the capacity of an accountability mechanism to achieve sustained results. We observed that in the **pre-implementation** phase of the mechanism, success factors included engagement of a multisectoral, diverse group of stakeholders and taking into account the context when establishing the criteria for monitoring and verifying adherence to them. Most of the studies in this review utilized a context analysis to inform their intervention design, but few achieved higher institutionalization. This may be because key stakeholders that were needed for the sustainability of the approaches were not sufficiently engaged in the accountability process. Where success was achieved, key ingredients for success were cited as: multidisciplinary and multi-stakeholder engagement, internal champions, and national advocates with the political capital to advocate for the approach.

This finding is supported by a recent review of accountability related to sexual and reproductive health and rights, *(9)* which found that accountability requires the engagement of multiple actors with a range of roles and responsibilities to be effective within an "accountability ecosystem". Such ecosystems have many influencing factors to consider when designing and implementing complex accountability interventions and strategies. In the design phase of an intervention, we also observe that a thorough understanding of the context, including power relationships between internal and external actors, is fundamental to developing and eventually institutionalizing accountability structures. A recent analysis found that government responsiveness to demands for improved accountability is influenced by a multitude of factors that shape bureaucratic decision-making, and is highly context specific, including structural factors such as resources, as well as willingness to act and the importance of separating and accommodating the underlying factors that can influence the effectiveness of an intervention's trajectory either towards or away from sustainability and institutionalization. *(80)*

Within the studies reviewed, the actors implicated in the professional accountability intervention tend to be dominated by facility-based health workers, and as noted above, important groups such as patients (who are usually the rights-holder but are hardly ever acknowledged as an actor within the mechanism). As a result, the importance of multi-stakeholder engagement in the process from the design to the impact may be overlooked. In the pre-implementation phase it is critical to involve all key potential power holders and decision makers that could eventually be the allies and advocates (change makers) needed to ensure a remedial action is taken, processes are sustained and eventually, policy and practice can one day change professional practice to institutionalize the accountability mechanism within the system. *(81)*

The **implementation phase** of an accountability mechanism aims to iteratively monitor, review and act on data collected about the issue. All of the studies in this review collected data to monitor system performance, usually through tools such as checklists, forms, scorecards, and tracking systems. The greater the transparency, criteria-based selection of the data collected, and external verification of the process, the greater the perceived authenticity of the resulting information. Once collected, review of the data to assess gaps and shortcomings in the accountability loop was generally the focus of the accountability process. Problems were identified in this step of the process and solutions proposed to remedy the problems. Some of the studies indicated that the review process led to the identification of a remedial action, which was then implemented to solve the identified problem. Of the 36 studies included in this review however, only 20 of the predominately clinical intervention studies (clinical audit and support supervision) reached the "act" stage of the implementation phase. It appears that the very process of collecting data and then reviewing it to identify solutions – without taking the added step of implementing a verifiable process for answerability – was seen as sufficient to elicit a professional accountability response. According to Brinkerhoff and others, accountability means being answerable to an agreed set of goals and objectives underpinned by clear roles and responsibilities. *(76)* They note that failure of those responsible to meet their agreed obligations should be met by formal or informal sanctions that are enforced. Failure to enforce sanctions against those found to be in violation of agreed roles and responsibilities diminishes accountability and trust in the system. *(77)* The lack of recourse embedded within the accountability processes put forward in the studied articles calls into question the collective understanding of professional accountability as a process by which those responsible are held to account for their action or inaction. Increased emphasis on **recourse and remedy** is urgently needed if accountability efforts are to add more than the past and current rhetoric on monitoring, analysis (review) and discussion of lessons learned and progress made, to truly achieve accountability on the SDGs for WCAH.

Monitor, review and act are the fundamental ingredients of the Universal Accountability Framework, but in this study we also considered: (a) pre-implementation design and stakeholder engagement, and (b) the post-implementation phase of the processes, to capture the life cycle of systemic change in a governance structure if the accountability mechanism is to be sustained by the system after the initial intervention stage is concluded. Only six studies recorded information that suggested there had been country-led, sustained changes to processes as a result of the mechanism, *(40,44,46–49)* and none indicated transformative

change such as changes to norms, despite calls for this within the EWEC strategy. The lack of remedy to problems identified in the implementation phase of the accountability intervention process results in short-lived or muted impact, at best. The six studies that did report some level of sustained change share a few common characteristics: all utilized an accreditation/certification system and benchmarking around accreditation and standard setting to elicit change in practice. The assessment services data or review of practice was done internally by senior staff and peers in a transparent manner with established criteria that all staff understood and agreed to adhere to. Identified problems were discussed by the service providers and changes were prioritised internally. Greater awareness of expected performance through in-depth discussion and training on quality measures, standards or protocols generated greater adherence and thus better performance. Improvement in turn increased staff motivation. Visible results, as measured through accreditation, and reductions in undesired effects and code violations (related to safety protocols), served as "rewards" for improved service delivery. Changes in services delivery standards and the workplan norms associated with them were seen as achievements that should be maintained by the workforce and managers alike. This in turn led to institutionalized changes in the workplace culture, which sustained the quality improvements over time.

The achievements reported in these studies were observed in HICs in specialized service settings, likely with a more homogeneous workforce. Challenges such as low salaries and other structural constraints often experienced in LMIC health service settings adversely affect staff motivation and capacity to change practice and must not be under-appreciated. Nonetheless, the factors that led to institutionalization of quality improvements in these settings are the very same factors appreciated in LMIC contexts, indicating that internally-led processes that are undertaken voluntarily and valued by staff, and are externally recognized by the system (and possibly the public at large) are key ingredients to establishing functional accountability systems in WCAH care settings. Sustaining improved processes and maintaining an accountability feedback loop require dedicated staff and managers. Costs associated with such processes should be covered from existing budget allocations for supervision, however when resources are constrained there will be a strong temptation to cut these budgets. Prioritising accountability to ensure quality of care will need to be defined and costed if they are to become part of the health service performance and quality improvement system.

There are some clear examples of how implementation of accountability mechanisms can improve professional performance, and even in some cases, health outcomes. Yet the majority of the evidence in this review and other recent reviews across the accountability domain, *(5, 9, 13)* has struggled to deliver evidence on 'what works' due to the complexity of the interventions and the appropriateness of the methods used to evaluate them. Randomized controlled trials (RCTs), as the gold standard for determining cause and effect, explore only a very small and highly-focused aspect of complex interventions and thus may not be wholly appropriate as a study design for such interventions. Some of the interventions studied through RCTs were quite weak due to this necessarily narrow investigative lens, which might be better captured through mixed methodologies or qualitative studies for such complex intervention pathways. These designed interventions thus suffer from a lack of quality and grounding of the intervention itself, and moreover, the appropriateness of the RCT methodology to capture the change process from the accountability intervention. Nevertheless, there have been many high quality studies, which tell a mixed story of 'what works' despite field experience from social accountability advocates that tell more positive results. *(80, 83)*

Some argue that this is because the model does not lend itself to study through traditional methods such as randomized controlled trials *(4,5)* as many of the effects of accountability are process changes rather than outcome oriented. Others see the need to contextualize the results, particularly related to social accountability. Also noted by many are the unintended effects of accountability efforts. *(81)* Bias in evaluating effects by implementers has also been a noted problem further limiting what is known from PBF interventions. *(80,82)* To build a strong body of evidence, short and long haul effects should be measured. Effects on practice norms and values such as changes in provider attitudes and practices, and even appreciation of the feedback loop and support structures sometimes embedded in professional accountability mechanisms, should also be considered as measures of success. Reliance on quality and

health outcomes only, may be underestimating the contribution of accountability efforts to change long haul effects. Theory based evaluations with mixed methods, realist evaluation, and other methods hold promise as they allow for both rigor and the exploration of the unknown or unexpected consequences. For example, it has become clear that for social accountability interventions, facilitation and information are needed for success. Yet only information provision, and not facilitation (i.e. who are the champions, how are they facilitated, the social action and engagement, etc.), can be readily assessed or "measured" through randomized controlled trials. *(81,82)*

The study of professional accountability needs to explore "how" and "why" some accountability mechanisms effect change in a given context, with specific supporting factors such as effective facilitation or the process. For professional accountability specifically, success must also capture the answerability of the accountability mechanism and process. Professional accountability relies less on advocacy and more specifically on the feedback loop or resources if the identified problem which is the focus of the intervention is redressed – the greater the enforceability, the stronger the effect. (10, 15) Researchers should consider how accountability mechanisms operate and in what context potential benefits may be seen to improve performance in the health sector.

Conclusion

This review indicates that a multidisciplinary approach to accountability in practice is essential to sustain improvements. Likewise, if we are to advance the field of accountability, there needs to be greater understanding of and clarity about how accountability mechanisms (rather than specific tools or approaches) should operate within a health systems improvement cycle, who should be engaged to lead and participate in such a process, and how success should be measured and ultimately sustained.

Greater appreciation of the context and consumers of the changes requested, as well as the interests and constraints of those required to make the changes, will be needed. A sound, evidence-based theory of change and preliminary political economy analysis in the design phase of any accountability intervention will assure the right actors are engaged from the start and that all stakeholders have a common vision of the change pathway. Many have written about the importance of understanding why change happens in some settings and not others; why some accountability approaches are effective and sustained though the majority are not; why government is responsive to social, professional/performance or even financial accountability efforts in some circumstances and not in others. (8,9,13,80,81) This review complements the findings of other recent studies and reviews, which indicate that while robust evidence is scarce, the underlying factors that influence the success of an accountability mechanism's contribution to change – particularly in quality of care – are shared. Those concerned with improving WCAH outcomes and health system performance via increased professional accountability can use these findings to increase accountability in practice. It will be important to ensure that implementation of such mechanisms however, is supported and sustained, so that accountability will one day be institutionalized and contribute to transforming service performance for improved women's, children's and adolescents' health. The achievement of global goals, and the EWEC Global Strategy depend upon it.

References

(1) Martin Hilber A, Blake C, Bohle LF, et al. Strengthening accountability for improved maternal and newborn health: A mapping of studies in Sub-Saharan Africa. Int. J. Gynecol. Obstet. 2016;135:345–357.

(2) Olafsdottir AE, Reidpath DD, Pokhrel S, et al. Health systems performance in sub-Saharan Africa: governance, outcome and equity. BMC Public Health. 2011;11:237.

(3) WHO, EWEC. Every Woman Every Child: A post-2015 vision. The third report of the independent Expert Review Group of the Commission on Information and Accountability for Women's and Children's Health. Geneva: World Health Organization; 2014.

(4) Bhutta ZA, Chopra M. Moving ahead: what will a renewed Countdown to 2030 for Women and Children look like? Lancet. 2016;387:2060–2062.

(5) EWEC. Unified Accountability Framework [website]. 2016 (https://www.everywomaneverychild.org/global-strategy/unified-accountability-framework/, accessed 18 October 2018).

(6) Cleary SM, Molyneux S, Gilson L. Resources, attitudes and culture: An understanding of the factors that influence the functioning of accountability mechanisms in primary health care settings. BMC Health Serv. Res. 2013;13:320.

(7) Lodenstein E, Dieleman M, Gerretson B, et al. Health provider responsiveness to social accountability initiatives in low- and middle-income countries: A realist review. Health Policy Plan. 2017;32:125–140.

(8) Fryatt R, Bennett S, Soucat A. Health sector governance: Should we be investing more? BMJ Glob. Heal. 2017;2:e000343.

(9) Molyneux S, Atela M, Angwenyi V, et al. Community accountability at peripheral health facilities: A review of the empirical literature and development of a conceptual framework. Health Policy Plan. 2012;27:541–554.

(10) Van Belle S, Boydell V, George AS, et al. Broadening understanding of accountability ecosystems in sexual and reproductive health and rights: A systematic review. PLoS One. 2018;13:e0196788.

(11) Brinkerhoff D. Taking account of accountability: A conceptual overview and strategic options. Washington D.C.: USAID; 2001.

(12) Mubyazi GM, Hutton G. Rhetoric and reality of community participation in health planning, resource allocation and service delivery: A review of the reviews, primary publications and grey literature. Rwanda J. Heal. Sci. 2012;1:51–65.

(13) Rifkin SB. Examining the links between community participation and health outcomes: A review of the literature. Health Policy Plan. 2014;4:29.

(14) Fox JA. Social accountability: What does the evidence really say? World Dev. 2015;72:346–361.

(15) Accountability in governance. World Bank; 2017 (https://siteresources.worldbank.org/PUBLICSECTORANDGOVERNANCE/Resources/AccountabilityGovernance.pdf, accessed 15 September 2017).

(16) Shaw CD. Toolkit for accreditation programs: Some issues in the design and redesign of external health care assessment and improvement systems. Melbourne: The International Society for Quality in Health Care; 2004.

(17) Ellis J. All inclusive benchmarking. J. Nurs. Manag. 2006;14:377–383.

(18) Ettorchi-Tardy A, Levif M, Michel P. Benchmarking: A method for continuous quality improvement in health. Healthc. Policy. 2012;7:e101–e119.

(19) National Institute for Health and Care Excellence. Glossary: Clinical audit; 2018 (https://www.nice.org.uk/Glossary?letter=C, accessed 15 September 2018).

(20) Training for mid-level managers (MLM): 4. Supportive supervision. Geneva: World Health Organization; 2008.

(21) Bosch-Capblanch X, Liaqat S, Garner P. Managerial supervision to improve primary health care in low- and middle-income countries (Review). Cochrane Database Syst. Rev. 2011;DOI: 10.1002/14651858.CD006413.pub2.

(22) Witter S, Fretheim A, Kessy FL, et al. Paying for performance to improve the delivery of health interventions in low- and middle-income countries. Cochrane Database Syst. Rev. 2012;2:DOI: 10.1002/14651858.CD007899.pub2.

(23) Bertone MP, Meessen B. Studying the link between institutions and health system performance: a framework and an illustration with the analysis of two performance-based financing schemes in Burundi. Health Policy Plan. 2013;28:847–857.

(24) Willcox ML, Nicholson BD, Price J, et al. Death audits and reviews for reducing maternal, perinatal and child mortality: Protocol for a Cochrane Review (Intervention). Cochrane Database Syst. Rev. 2018;DOI: 10.1002/14651858.CD012982.

(25) Kongnyuy EJ, Uthman OA. Use of criterion-based clinical audit to improve the quality of obstetric care: A systematic review. Acta Obs. Gynecol Scand. 2010;88:873–881.

(26) Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: Effects on professional practice and healthcare outcomes (Review). Cochrane Database Syst. Rev. 2012;DOI: 10.1002/14651858.CD000259.pub3.

(27) Gorter A, Ir P, Meessen B. Evidence Review, Results-Based Financing of Maternal and Newborn Health Care in Low- and Lower-middle-Income Countries. Bonn: German Federal Ministry for Economic Cooperation and Development (BMZ) through the sector project PROFILE at GIZ – Deutsche Gesellschaft für Internationale Zusammenarbeit; 2013.

(28) Das A, Gopalan SS, Chandramohan D. Effect of pay for performance to improve quality of maternal and child care in low- and middle-income countries: a systematic review. BMC Public Health. 2016;16:321.

(29) Suthar AB, Nagata JM, Nsanzimana S, et al. Performance-based financing for improving HIV/AIDS service delivery: a systematic review. BMC Health Serv. Res. 2017;16:doi: 10.1186/s12913-016-1962-9.

(30) Scott A, Sivey P, Ait Ouakrim D, et al. The effect of financial incentives on the quality of health care provided by primary care physicians. Cochrane Database Syst. Rev. 2011;9:DOI: 10.1002/14651858.CD008451.pub2.

(31) Eijkenaar F, Emmert M, Scheppach M, et al. Effects of pay for performance in health care: a systematic review of systematic reviews. Health Policy (New. York). 2013;110:115–130.

(32) Gopalan SS, Mutasa R, Friedman J, et al. Health sector demand-side financial incentives in low- and middle-income countries: a systematic review on demand- and supply-side effects. Soc. Sci. Med. 2014;100:72–83.

(33) Turcotte-Tremblay A-M, Spagnolo J, De Allegri M, et al. Does performance-based financing increase value for money in low- and middle- income countries? A systematic review. Health Econ. Rev. 2016;6:30.

(34) Gergen J, Josephson E, Coe M, et al. Quality of Care in Performance-Based Financing: How It Is Incorporated in 32 Programs Across 28 Countries. Glob. Heal. Sci. Pract. 2017;5:90–107.

(35) Munar W, Snilstveit B, Stevenson J, et al. Evidence gap map of performance measurement and management in primary care delivery systems in low- and middle-income countries – Study protocol. Gates Open Res. 2018;2:27.

(36) Renmans D, Holvoet N, Criel B, et al. Performance-based financing: The same is different. Health Policy Plan. 2017;32:860–868.

(37) Kambala C, Lohmann J, Mazalale J, et al. Perceptions of quality across the maternal care continuum in the context of a health financing intervention: Evidence from a mixed methods study in rural Malawi. BMC Health Serv. Res. 2017;17:392.

(38) Puhan MA, Schunemann HJ, Murad MH, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ. 2014;349:g5630.

(39) Berger ER, Wang CE, Kaufman CS, et al. National accreditation program for breast centers demonstrates improved compliance with post-mastectomy radiation therapy quality measure. J. Am. Coll. Surg. 2017;224:236–244.

(40) Peterson TH, Teman SF, Connors RH. A safety culture transformation: Its effects at a children's hospital. J. Patient Saf. 2012;8:125–130.

(41) Banerjee SK, Andersen KL, Navin D, et al. Expanding availability of safe abortion services through private sector accreditation: a case study of the Yukti Yojana program in Bihar, India. Reprod. Health. 2015;12:104.

(42) Mekory TM, Bahat H, Bar-Oz B, et al. The proportion of errors in medical prescriptions and their executions among hospitalized children before and during accreditation. Int. J. Qual. Heal. Care. 2017;29:366–370.

(43) Margeirsdottir H, Larsen J, Kummernes S, et al. The establishment of a new national network leads to quality improvement in childhood diabetes: Implementation of the ISPAD Guidelines. Pediatr. Diabetes. 2010;11:88–95.

(44) Adams M, Heohre TC, Bucher HU. The swiss neonatal quality cycle, a monitor for clinical performance and tool for quality improvement. BMC Pediatr. 2013;13:152.

(45) Brucker SY, Wallwiener M, Kreienberg R, et al. Optimizing the Quality of Breast Cancer Care at Certified German Breast Centers. Strahlenther Onkol. 2011;187:89–99.

(46) van Dam P, Tomatis M, Marotti L, et al. Time trends (2006-2015) of quality indicators in EUSOMA-certified breast centres. Eur. J. Cancer Care (Engl). 2017;85:15–22.

(47) Kilsdonk MJ, van Dijk BA, Otter R, et al. Regional variation in breast cancer treatment in the Netherlands and the role of external peer review: a cohort study comprising 63,516 women. BMC Cancer. 2014;14:596.

(48) Harrison K, Peek J, Chapman M, et al. Continuous improvement in national ART standards by the RTAC accreditation system in Australia and New Zealand. Aust. New Zeal. J. Obstet. Gynaecol. 2017;57:49–51.

(49) Rizzo M, Bumpers H, Okoli J, et al. Improving on national quality indicators of breast cancer care in a large public hospital as a means to decrease disparities for African American women. Ann. Surg. Oncol. 2011;18:34–39.

(50) Yilla M, Nam SL, Adeyemo A, et al. Using scorecards to achieve facility improvements for maternal and newborn health. Int. J. Gynecol. Obstet. 2014;127:108–112.

(51) Veerbeek L, van der Geest L, Wouters M, et al. Enhancing the quality of care for patients with breast cancer: Seven years of experience with a Dutch auditing system. EJSO. 2011;37:714–718.

(52) Bailey P, Binh H, Bang H. Promoting accountability in obstetric care: Use of criteria-based audit in Viet Nam. Glob. Public Health. 2010;5:62–74.

(53) Dupont C, Deneux-Tharaux C, Touzet S, et al. Clinical audit: a useful tool for reducing severe postpartum haemorrhages? Int. J. Qual. Heal. Care. 2011;23:583–589.

(54) Hunyinbo K, Fawole A, Sotiloye O, et al. Evaluation of criteria-based clinical audit in improving quality of obstetric care in a developing country hospital. Afr. J. Reprod. Health. 2008;12:59–70.

(55) Kidanto HL, Wangwe P, Kilewo CD, et al. Improved quality of management of eclampsia patients through criteria based audit at Muhimbili National Hospital, Dar es Salaam, Tanzania. Bridging the quality gap. BMC Pregnancy Childbirth. 2012;12:134.

(56) Ngwenya S. Reducing fresh full term intrapartum stillbirths through leadership and accountability in a low-resource setting, Mpilo Central Hospital, Bulawayo, Zimbabwe. BMC Res. Notes. 2017;10:246.

(57) Pirkle CM, Dumont A, Traoré M, et al. Effect of a facility-based multifaceted intervention on the quality of obstetrical care: a cluster randomized controlled trial in Mali and Senegal. BMC Pregnancy Childbirth. 2013;13:24.

(58) van Dillen J, Lim F, van Rijssel E. Introducing caesarean section audit in a regional teaching hospital in the Netherlands. Eur. J. Obstet. Gynecol. Reprod. Biol. 2008;139:151–156.

(59) Deneux-Tharaux C, Dupont C, Colin C, et al. Multifaceted intervention to decrease the rate of severe postpartum haemorrhage: the PITHAGORE6 cluster-randomised controlled trial. BJOG. 2010;117:1278–1287.

(60) Bayley O, Chapota H, Kainja E, et al. Community-linked maternal death review (CLMDR) to measure and prevent maternal mortality: A pilot study in rural Malawi. BMJ Open. 2015;5:e007753.

(61) Dumont A, Fournier P, Fraser W, et al. QUARITE (quality of care, risk management and technology in obstetrics): A cluster-randomized trial of a multifaceted intervention to improve emergency obstetric care in Senegal adn Mali. Trials. 2009;10:85.

(62) Gebrehiwot Y, Tewolde BT. Improving maternity care in Ethiopia through facility based review of maternal deaths and near misses. Int. J. Gynecol. Obstet. 2014;127:S29–S34.

(63) Hofman JJ, Mohammed H. Experiences with facility-based maternal death reviews in northern Nigeria. Int. J. Gynecol. Obstet. 2014;126:111–114.

(64) Kongnyuy E, Leigh B, van den Broek N. Effect of audit and feedback on the availability, utilisation and quality of emergency obstetric care in three districts in Malawi. Women and Birth. 2008;21:149–155.

(65) Nakibuuka V, Okong P, Waiswa P, et al. Perinatal death audits in a peri-urban hospital in Kampala, Uganda. Afr. Health Sci. 2012;4:435–442.

(66) Okonofua F, Imosemi D, Igboin B, et al. Maternal death review and outcomes: An assessment in Lagos State, Nigeria. PLoS One. 2017;12:e0188392.

(67) Sorensen BL, Elsass P, Bruun Nielsen B, et al. Substandard emergency obstetric care – a confidential enquiry into maternal deaths at a regional hospital in Tanzania. Trop. Med. Int. Heal. 2010;15:894–900.

(68) Zongo A, Dumont A, Fournier P, et al. Effect of maternal death reviews and training on maternal mortality among cesarean delivery: post-hoc analysis of a cluster-randomized controlled trial. Eur. J. Obstet. Gynecol. Reprod. Biol. 2015;185:174–180.

(69) Bouvier-Colle M-H, Saucedo M, Deneux-Tharaux C. L'enquête confidentielle française sur les morts maternelles, 1996-2006 : Quelles conséquences pour les soins en obstétrique ? J. Gynécologie Obs. Biol. la Reprod. 2011;40:87–102.

(70) Moodley J, Pattinson R, Fawcus S, et al. The confidential enquiry into maternal deaths in South Africa: a case study. BJOG. 2014;121:53–60.

(71) Paily V, Ambujam K, Rajasekharan Nair V, et al. Confidential review of maternal deaths in Kerala: a country case study. BJOG. 2014;121:61–66.

(72) Jenkins KJ, Castañeda AR, Cherian K, et al. Reducing mortality and infections after congenital heart surgery in the developing world. Pediatrics. 2014;134:e1422–e1430.

(73) McPhail GL, Weiland J, Acton JD, et al. Improving evidence-based care in cystic fibrosis through quality improvement. Arch. Pediatr. Adolesc. Med. 2010;164:957–960.

(74) Reynolds HW, Toroitich-Ruto C, Nasution M, et al. Effectiveness of training supervisors to improve reproductive health quality of care: a cluster-randomized trial in Kenya. Health Policy Plan. 2008;23:56–66.

(75) Brinkerhoff D. Accountability and Health Systems: Overview, Framework, and Strategies. Bethesda: The Partners for Health Reformplus Project, Abt Associates Inc; 2003.

(76) Brinkerhoff DW. Accountability and health systems: Toward conceptual clarity and policy relevance. Health Policy Plan. 2004;19:371–379.

(77) Suter E, Mallinson S. Accountability for coordinated/integrated health services delivery: Working paper. Copenhagen: WHO Regional Office for Europe; 2015.

(78) Graham W, Wagaarachchi P, Penney G, et al. Criteria for clinical audit of the quality of hospital-based obstetric care in developing countries. Bull. World Health Organization. 2000;78:614–620.

(79) Rockers PC, Kruk ME, Laugesen MJ. Perceptions of the health system and public trust in government in low and middle-income countries: Evidence from the World Health Surveys. J. Health Polit. Policy Law. 2012;37:405–437.

(80) Joshi A, McCluskey R. The art of 'bureaucraft': Why and how bureaucrats respond to citizen voice, Making All Voices Count Research Briefing. Brighton: Institute of Development Studies; 2018.

(81) Schaaf M, Topp SM, Ngulube M. From favours to entitlements: community voice and action and health service quality in Zambia. Health Policy Plan. 2017;32:847–859.

(82) Westhorp G, Walker B, Rogers P, et al. Enhancing community accountability, empowerment and education outcomes in low and middle-income countries: A realist review. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London; 2014.

(83) Gaventa J, Barrett G. So what difference does it make? Mapping the outcomes of citizen engagement. IDS working paper 347. Brighton; 2010. Report No.: 347.

Annex: Search terms

The exact search terms varied according to the conventions used by each search engine, as follows:

PubMed

The following filters were applied:

- Article types: Journal article
- Publication dates: 10 years
- Species: humans
- Languages: English, French, Spanish, Portuguese

SEARCH 1

(human resources OR doctors OR nurses OR midwives OR hospital OR clinic OR health centre OR hygiene OR sanitation) AND (women OR reproductive OR maternal OR pregnancy OR birth OR breast OR cervical OR adolescents OR teenagers OR child OR neonatal OR newborn OR infant) AND (accountability OR accreditation OR (performance AND appraisal) OR (performance AND audit) OR benchmarking OR grievance OR confidential enquiry OR confidential inquiry OR league table OR licence to practise OR oversight OR redress OR sanctions OR scorecard OR dashboard OR transparency)

SEARCH 2

(accountability AND health system performance)

SEARCH 3

((((maternal death review*[Title/Abstract]) OR perinatal death review*[Title/Abstract]) OR maternal death surveillance[Title/Abstract]) AND ("last 10 years"[PDat] AND Humans[Mesh] AND (English[lang] OR French[lang] OR Portuguese[lang] OR Spanish[lang])))

SEARCH 4

((health facility committee*[Title/Abstract]) OR performance-based financ*[Title/Abstract]) OR results-based financ*[Title/Abstract]

CINAHL/EBSCO

The following filters were applied:

- Published between Jan 2008 and May 2018
- Human
- Exclude MEDLINE records
- Research articles
- Academic journals

Search terms: TX ((human resources OR doctors OR nurses OR midwives OR hospital OR clinic OR health centre OR hygiene OR sanitation)) AND TX ((women OR reproductive OR maternal OR pregnancy OR birth OR breast OR cervical OR adolescents OR teenagers OR child OR neonatal OR newborn OR infant)) AND TX ((accountability OR accreditation OR (performance AND appraisal) OR (performance AND audit) OR benchmarking OR grievance OR confidential enquiry OR confidential inquiry OR league table OR licence to practise OR oversight OR redress OR sanctions OR scorecard OR dashboard OR transparency))

Web of Science

Search terms: health AND (health system OR human resources OR doctors OR nurses OR midwives OR hospital OR clinic OR health centre OR hygiene OR sanitation) AND (women OR reproductive OR maternal OR pregnancy OR birth OR breast OR cervical OR adolescents OR teenagers OR child OR neonatal OR newborn OR infant) AND (accountability OR accreditation OR (performance AND appraisal) OR (performance AND audit) OR benchmarking OR grievance OR confidential enquiry OR confidential inquiry OR league table OR licence to practise OR oversight OR redress OR sanctions OR scorecard OR dashboard OR transparency)

Refined by: DOCUMENT TYPES: (ARTICLE) AND WEB OF SCIENCE CATEGORIES: (EDUCATION SPECIAL OR HEALTH CARE SCIENCES SERVICES OR PEDIATRICS OR OBSTETRICS GYNECOLOGY OR SOCIAL ISSUES OR HEALTH POLICY SERVICES OR ANTHROPOLOGY OR NURSING OR ONCOLOGY OR SOCIAL SCIENCES INTERDISCIPLINARY OR DEMOGRAPHY OR PSYCHIATRY OR INFECTIOUS DISEASES OR PRIMARY HEALTH CARE OR SOCIAL SCIENCES BIOMEDICAL OR MULTIDISCIPLINARY SCIENCES OR BEHAVIORAL SCIENCES OR SOCIAL WORK OR EDUCATION SCIENTIFIC DISCIPLINES OR WOMEN S STUDIES OR FAMILY STUDIES OR NUTRITION DIETETICS OR SOCIOLOGY OR EDUCATION EDUCATIONAL RESEARCH OR PUBLIC ADMINISTRATION OR TROPICAL MEDICINE OR INTERNATIONAL RELATIONS) AND [excluding] WEB OF SCIENCE CATEGORIES: (HEALTH CARE SCIENCES SERVICES OR PEDIATRICS OR PUBLIC ENVIRONMENTAL OCCUPATIONAL HEALTH OR ONCOLOGY OR PSYCHIATRY OR INFECTIOUS DISEASES OR SOCIAL SCIENCES BIOMEDICAL OR SOCIAL WORK OR FAMILY STUDIES OR MEDICINE GENERAL INTERNAL OR IMMUNOLOGY OR CLINICAL NEUROLOGY OR MEDICAL INFORMATICS OR SURGERY) AND [excluding] WEB OF SCIENCE CATEGORIES: (REPRODUCTIVE BIOLOGY OR INDUSTRIAL RELATIONS LABOR OR UROLOGY NEPHROLOGY OR CRITICAL CARE MEDICINE OR GERIATRICS GERONTOLOGY OR HISTORY PHILOSOPHY OF SCIENCE OR INTERNATIONAL RELATIONS OR LAW OR PARASITOLOGY OR POLITICAL SCIENCE OR PSYCHOLOGY EDUCATIONAL OR SPORT SCIENCES OR AGRICULTURE MULTIDISCIPLINARY OR ANESTHESIOLOGY OR ANTHROPOLOGY OR AREA STUDIES OR BEHAVIORAL SCIENCES OR ENDOCRINOLOGY METABOLISM OR FOOD SCIENCE TECHNOLOGY OR GEOGRAPHY OR HISTORY OR HISTORY OF SOCIAL SCIENCES OR HOSPITALITY LEISURE SPORT TOURISM OR ORTHOPEDICS OR REHABILITATION OR RELIGION) AND [excluding] WEB OF SCIENCE CATEGORIES: (SOCIOLOGY OR EDUCATION SCIENTIFIC DISCIPLINES OR EDUCATION SPECIAL)

Time span: 2008-2018. Indexes: SCI-EXPANDED, SSCI, A&HCI, ESCI, CCR-EXPANDED, IC.

Cochrane Library

The following filters were applied:

- Between 2008 and 2018
- Trials only

Search terms: Health and accountability

Virtual Regional Portal for PAHO

The following filters were applied:

- 2008 or later
- LILACS
- English, Portuguese, Spanish
- Articles

Search terms: Responsabilidade e Sistema de saúde

The Partnership
for Maternal, Newborn
& Child Health

Professional accountability for women's, children's and adolescents' health: what mechanisms and processes are used, what works?

A systematic literature review