

# **Community Monitoring Interventions to Curb Corruption and Increase Access and Quality in Service Delivery: A Systematic Review**

## **1. Introduction**

Corruption and inefficient allocation of resources in service delivery are widespread in low and middle income countries (Pande & Olken, 2011). There is increasing evidence that corruption holds back countries' economic development and erodes their citizens' quality of life (Mauro, 1995; Svensson, 2005; Singer, 2013). Millions of people around the world encounter administrative corruption in their daily interactions with public services (Transparency International, 2013). The World Bank Institute estimates that total bribes in a year amount to about one trillion USD (Rose-Ackerman, 2004), making corruption account for around three per cent of world GDP (Svensson, 2005). Bribes are used to influence the actions of public officials, either to performed their duties, distort the duties or to prevent them from performing their duties.

In many countries, corruption is widespread throughout the public sector, not only among high level public officials. Administrative corruption imposes a heavy burden on citizens' and firms' time and resources. For example, Olken and Barron (2009) estimate that 13 per cent of the cost of a truck driver's trip in Indonesia is allocated to pay bribes to police officials that they encounter on their journey. In Uganda, bribes represent eight per cent of firms' production costs (Svensson, 2003). Corruption creates discontent with public services, undermines trust in public institutions (Sacks & Larizza, 2012; Singer, 2013), and stifles business growth and investment.

Often, it is the poor and the vulnerable who suffer the most from public sector corruption (Olken, 2006; Sukhtankar, 2011). A landmark study in Uganda found that only 13 per cent of the public funds that the central government had assigned to the

school system reached the intended destination (Reinikka & Svensson, 2004, 2005, 2011). In India, the lack of monitoring and accountability has resulted in high levels of public sector absenteeism, with one quarter of all the teachers in public schools and more than a third of nurses and doctors being absent from their duties (Chaudhury *et al.*, 2006).

Community Monitoring Interventions (CMI) aim to address such issues and have become common in recent years. Such programmes seek to involve communities in the monitoring of public service providers to increase their accountability to users. However, their effectiveness in reducing corruption and improving access and quality of services remain unclear. A number of well-known studies of CMIs purport to demonstrate the effectiveness of CMIs.

Over the last decade the World Bank alone has channelled USD 85 billion to local participatory development (Mansuri & Rao, 2012). The United Nations has set increasing citizen participation as its main strategy to achieve good governance and human rights (UN, 2008). Non-governmental organizations (NGOs) with a focus on increasing government accountability through citizen participation continue to expand around the globe, managing increasing amounts of resources.<sup>i</sup>

Reports from major donors acknowledge that the lack of systematic evidence limits our ability to make precise claims regarding the relationship between CMIs, corruption and service delivery outcomes (Brinkerhoff & Azfar, 2008). There are a number of non-systematic literature reviews on community interventions or interventions for increasing citizen engagement in service delivery, e.g. Devarajan *et al.* (2011), Gaventa and Barret (2012), Mansuri and Rao (2012), Pande and Olken (2011), and Ringold *et al.* (2012). This is the first systematic review of the evidence on CMIs. We used systematic methods to search, appraise and synthesise evidence,

reaching lessons for policy transparently. Related systematic reviews have examined specific interventions aimed at reducing corruption (Hanna *et al.*, 2011) or improving social cohesion (King *et al.*, 2010).

The paper is structured as follows. In section 2 we present the intervention background and the theory of change. Section 3 presents objectives and the review approach. Section 4 presents search results, study descriptives and bias assessment. Section 5 gives results for primary and secondary outcomes. Finally, Sections 6 and 7 provide a discussion and conclusions for policy, programmes and research.

## **2. Community monitoring interventions**

The idea that community members have incentives to monitor providers and demand better services (Stiglitz, 2002) led practitioners to believe that allowing communities to have monitoring power over providers could be beneficial for improving service delivery and reducing corruption in both the short and long term. In the short term, it could improve outcomes by identifying pockets of corruption and inefficiency in service delivery. In the long term it may contribute to changes in political norms and to establishing a transparent and accessible channel of communication for the community to provide feedback to providers and politicians on a regular basis.

Over the last two decades programs aimed at encouraging community monitoring have been introduced in Albania, Argentina, Brazil, Cambodia, Cameroon, Colombia, Kenya, India, Indonesia, Malawi, Philippines, South Africa, and Uganda, among others (Reinikka & Svensson, 2004, 2005, 2011; Pan & Christiaensen, 2012; Tosi, 2010; Ferraz, Finan & Moreira, 2012; Capuno & Garcia, 2010; Ringold *et al.*, 2012).

Community monitoring interventions (CMIs), often referred to as social accountability mechanisms, are approaches where the community is given the opportunity to participate in the process of monitoring service delivery, where monitoring means being able to observe and assess providers' performance and provide feedback to providers and politicians. The Association for the Empowerment of Workers and Farmers in India was the first organization to introduce a social accountability initiative, through social audits in the early 1990s (Maru, 2010).

The four major categories of such interventions are information campaigns, scorecards and citizen report cards, social audits, and grievance redress mechanisms (Table 1). These four categories share two common elements: a clear objective of reducing corruption and improving service delivery; and, using encouragement of the community to monitor service delivery as a key intervention instrument.

*Table 1: Interventions to Increase Civic Participation in Monitoring*

<i>Intervention</i>	<i>Description</i>
<b>Information campaign</b>	These are efforts to inform citizens about their rights to services, quality standards, and performance campaigns. In particular, it can include information on the importance of the service, on providers' performance, and on how to monitor providers.
<b>Scorecards &amp; citizen report cards</b>	These involve quantitative surveys that assess users' satisfaction and experiences with various dimensions of service delivery. It often involves a meeting between the recipients of services and providers to discuss the findings of the survey and to develop a follow-up plan (Ringold <i>et al.</i> , 2012).
<b>Social audit</b>	Social audits allow citizens receiving a specific service to examine and cross-check the information the provider makes available against information collected from users of the service (Ringold <i>et al.</i> , 2012).
<b>Grievance redress mechanisms</b>	These are mechanisms that provide citizens with opportunities to use information redress to influence service delivery and give feedback on government programs and services, mechanisms including complaint hotlines, informal dispute resolution mechanisms, and courts (Ringold <i>et al.</i> , 2012).

Information campaigns are one of the most common interventions to encourage participation and interest in service delivery monitoring. They usually involve provision of information on the benefits of the service to be delivered (health, education, police, and so on) and the current state of the service in the community. The information could be provided door to door, in public gatherings aided by local leaders, through radio, newspapers or other means (e.g. Kefeer and Khemani, 2011). Information campaigns can also include information on how to monitor providers (e.g. Banerjee *et al.*, 2010).

Scorecards, often referred to as citizen report cards, encourage participation in monitoring service delivery by giving citizens a voice to demand better services. An important difference between information campaigns and scorecards is that the latter can include an interaction between citizens and providers, while the former does not include a forum for such interaction. For example, in Uganda, NGOs facilitated village and service provider staff meetings in which community members discussed information on the status of health service delivery, and were encouraged to develop plans identifying steps providers should take to improve service delivery (Björkman and Svensson, 2009).

Social audits involve interactions not only between citizens and providers, but also with politicians, as for instance in Colombia's Citizens Visible Audit (CVA) program which worked with public infrastructure projects in water and sanitation infrastructure, schools and hospitals (Molina, 2013b). A social audit involves: dissemination of information through radio, newspapers and local TV in the neighbourhoods where the project is due to take place; a public forum where citizens are introduced to the project and told about their rights and entitlements, including the activities they can do to monitor the project and the responsibilities of the executing firm; establishment and training of a group of beneficiaries to carry out community monitoring activities; periodic public meetings, bringing together local authorities, neighbours, and representatives from the firm carrying out the specific project, to discuss the state of the project and enable community members to voice concerns and the firm and local government to make commitments to solve the problems; regular monitoring of the project by the beneficiary group and collection of information on whether commitments are being honoured; presentation of the finalised project to the community before making the final payment to the executing firm, and sharing of the audit results with all interested and concerned stakeholders.

Social audits can also give citizens the power to make decisions about the project.

An example of a CMI where citizens had decision making power is the Kecamatan Development Program (KDP) in Indonesia, a community-driven development program that funded projects in about 15,000 villages each year (Olken, 2007).

Funds were paid to village 'implementation teams' in three instalments. To receive the second and third payments, the teams had to make accountability reports at an open village meeting.

Grievance redress mechanisms (GRMs) provide people with opportunities to use information to influence service delivery. GRMs capture different mechanisms that provide citizens with opportunities to use information redress to influence service delivery and give feedback on government programs and services. Such mechanisms include complaint hotlines, informal dispute resolution mechanisms, and courts (Ringold *et al.*, 2012).

We developed a stylized causal chain (Figure 1), which presents the processes involved in a typical CMI and the mechanisms through which the program is expected to have an impact on corruption and service delivery outcomes, where service delivery is defined as the process through which basic services, such as education, health, and security are delivered to communities.<sup>ii</sup> The causal chain suggests:

- CMIs will increase the quantity and the degree to which citizens are involved in monitoring service providers.
- As a result of the CMIs, politicians and providers will exert more effort and improve their performance in relation to service delivery.
- CMIs will reduce the probability of corruption.
- CMIs will improve access and quality of the service provided.

A typical CMI begins by attempting to make the project or service that it aims to monitor salient in the community. This is usually done through a communication campaign (building block 1) using as many mediums as possible, such as radio, newspapers, door to door campaigns, and local TV. The campaign's primary objective is to increase citizen knowledge of (a) the performance of the service to be monitored and/or (b) the importance of the service or project for the community.

Equipped with this information, citizens can engage in different activities. For instance, they might change their private actions, or contact fellow community members to collectively pressure providers and politicians to improve the quality of the service through monitoring activities (building block 2). To encourage citizens to monitor service providers, CMIs usually include activities to build the capacity of beneficiaries to monitor providers.

Empowered with information, citizens are expected to solve the collective action problem and invest their time and effort to participate in monitoring service delivery (building block 3). Participation in monitoring activities could take many forms, depending on the specific CMI. For instance, social audits have public forums and scorecards and can include meetings between providers and citizens.



As an organised group, citizens can take turns to visit the place where the service or project takes place, such as a school, construction site or hospital, and collect information on its problems, for example absenteeism, use of low quality inputs in the construction process, unresponsive front-line providers. Citizens can then contact providers (building block 6) and/or elected officials (building block 4) to file complains about the service and provide information on the specific problems the service is facing. In addition, citizens are expected to share the information collected by monitoring providers with their fellow neighbours that did not take part in monitoring activities (building block 5), to increase visibility of the community monitoring intervention and put pressure on providers and politicians.

Citizens' participation in the program may reduce the cost of monitoring front-line providers for politicians and managers. Citizens' monitoring activities also increase both visibility and citizens' ability to recognize whether elected officials are making an effort to reduce corruption and improve service delivery. As a result, there may be a greater incentive for politicians and policymakers to achieve better results and to put more pressure on providers to improve service delivery (building block 7). The threat of formal sanctions by politicians and/or informal sanctions by citizens is assumed to motivate service providers into exerting greater effort.

Many of these processes are mediated by local norms and context. Participation in the CMI will be influenced by the strength of the community to act collectively. For example, communities with a history of grassroots participation are expected to organise more rapidly and more efficiently (Björkman & Svensson, 2010). History can play an important role here. For example, in parts of Africa, the history of slave trade left an imprint in cultural norms and beliefs which arguably diminished the trust among fellow citizens and reduced the strength of the community to act collectively (Nunn & Wantchekon, 2011). The independence and strength of the local media is

also assumed to impact upon the visibility of the project (Reinikka & Svensson, 2005; Ringold *et al.*, 2012). Finally, this is a dynamic process, which makes understanding the specific history of service delivery, citizen engagement and political accountability in the community where the intervention took place crucial (Grandvoinnet *et al.*, 2015).

There are also some features that are specific to each intervention. Scorecards have an added accountability mechanism through which citizens meet with service providers to discuss how to improve the service. This face-to-face interaction introduces intrinsic motivation arguments for the service providers, which may contribute to improving their performance. This will be moderated by whether it is credible for a given community to establish an informal system of rewards and sanctions. Additionally, the meeting could result in new ideas for providers and citizens on how to use and manage the service in a more efficient way.

Social audits contain an additional component in the form of public forums that allow citizens to make their voice heard by local officials and providers, simultaneously reducing the time and effort citizens would need to invest to get an appointment with officials, costs for central government representatives to be heard by local officials, and costs of local officials to take actions to solve problems that arise during the implementation of the projects.

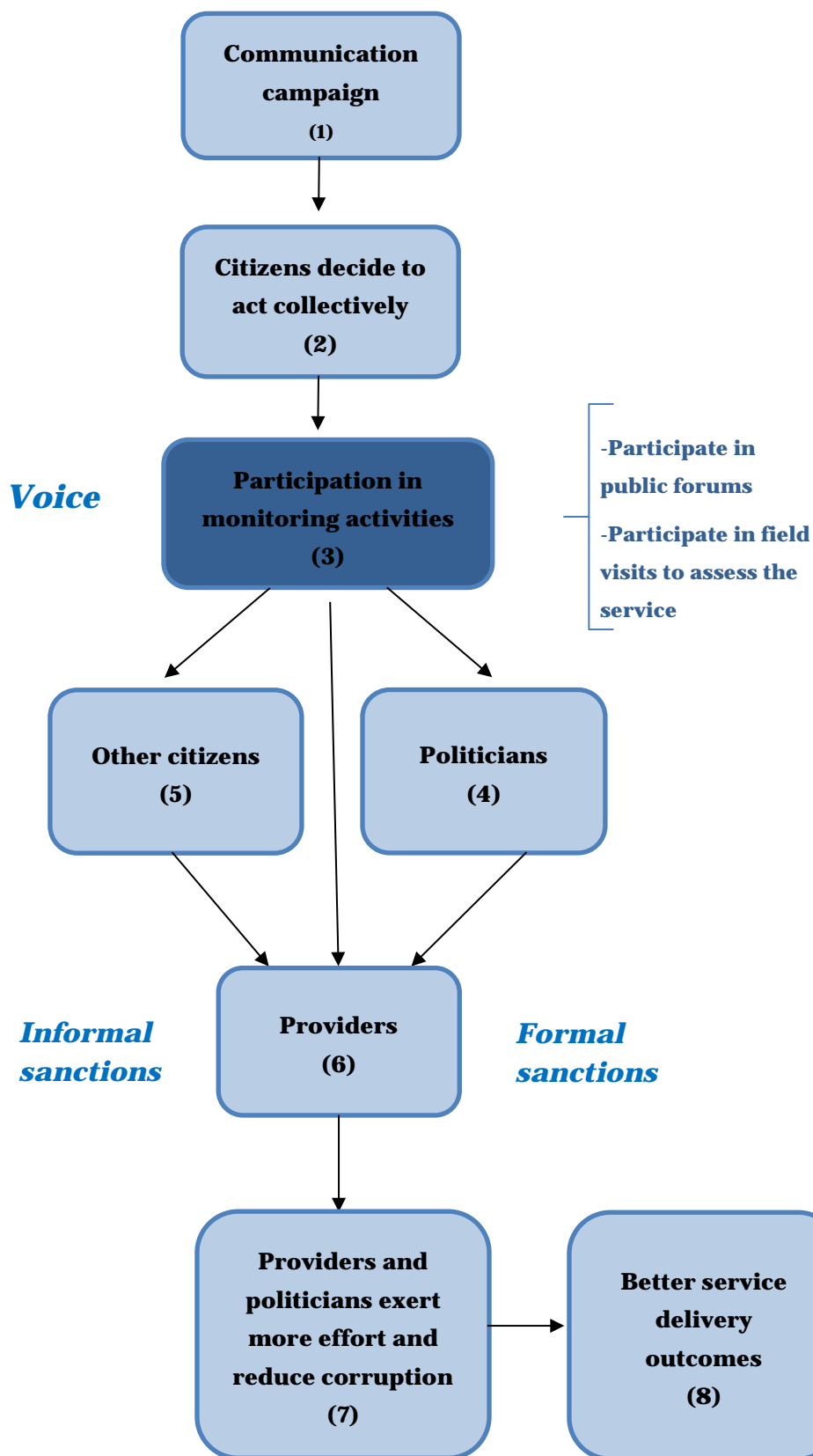
Several assumptions underlie the causal chain. Citizens need to participate in monitoring activities and politicians and providers need to be accountable. For citizens to participate, they need to have adequate information on how to monitor the project, be able to pay the opportunity cost of participation and coordinate their actions to monitor the project. Finally, citizens should believe the program has the potential to be successful, pay attention and be able to understand the information

provided. Providers and politicians need to gain popularity, increased salary and/or social recognition, obtain re-election or avoid social disapproval or an indictment. Whether or not these assumptions are met can affect citizens' decision on whether to monitor government activity and the governments' willingness to facilitate citizen engagement and become more accountable.

For example, one potential concern is that citizens will fail to participate in monitoring activities (building block 3). In particular, if community monitoring activities are not carried out, or carried out by only a few citizens, their ability to uncover problems and put pressure on the government to provide accountability can be significantly reduced.

Another potential concern is that politicians and providers may not be accountable to their citizens (building block 4 and 6). For example, if the community is not needed for a politician to stay in power, the politician's performance may not improve as a result of the CMI. In addition, in communities where providers are not responsive to politicians, CMIs will only be effective if they are able to change provider behaviour. Communities often have low state capacity, which limits the ability of governments to monitor providers (Besley & Persson, 2011). If this is the case, putting pressure on the government is likely to be ineffective and only competition or informal sanctions from the community may have an effect on providers' performance.

Figure 1: Causal Chain for Community Monitoring Interventions



Citizens need to participate in monitoring activities and politicians and providers need to be accountable. For citizens to participate, they need to have adequate information

on how to monitor the project, be able to pay the opportunity cost of participation and coordinate their actions to monitor the project. Finally, citizens pay attention to the information, be able to understand the information provided and believe the programme has the potential to be successful. Providers and politicians need to gain popularity, increased salary and/or social recognition, obtain re-election or avoid social disapproval or an indictment. If these assumptions are not met, the underlying programme theory of the CMI breaks down and this may prevent CMIs from having an impact on service delivery outcomes. In particular, whether or not they hold true can affect citizens' decision on whether to monitor government activity and the governments' willingness to facilitate citizen engagement and become more accountable. Below we present the bottlenecks as well as the empirical implications.

### **Civic participation failure**

One potential concern with CMIs is that citizens will fail to participate in monitoring activities (building block 3). We have identified six potential bottlenecks<sup>iii</sup> that could prevent citizens from participating in monitoring activities, which in turn reduces the potential impact of the programme (see Table 2). In particular, if community monitoring activities are not carried out, or carried out by only a few citizens, their ability to uncover problems and put pressure on the government to provide accountability can be significantly reduced.

Table 2: Bottlenecks preventing citizens from participating in monitoring activities

<b>Bottleneck</b>	<b>Description</b>	<b>Empirical Implications</b>
Information Gaps	Scholars and policymakers have long argued that programmes often fell short of their expectations because of information problems (Ringold <i>et al.</i> ,	<ul style="list-style-type: none"> <li>If the information is not properly disseminated, citizens will not</li> </ul>

	<p>2012). In the case of the CMIs there are two important potential deficiencies: (a) the information may not have been properly disseminated (building block 1), and/or (b) information on how to monitor the project was either not provided or not understood by the citizens (building block 2).</p>	<p>participate in monitoring activities</p> <ul style="list-style-type: none"> <li>• Citizens' probability of participation in monitoring activities will be a function of how well they understand how to monitor providers.</li> </ul>
<p>Lack of Attention Span or Rational Inattention</p>	<p>Even if information is provided, it may fail to have the anticipated outcome. A factor that conditions its success is what information is to be disclosed (content), and how it is to be presented (vehicle). In the case of CMIs, citizens' lack of attention span might prevent them from absorbing the information provided by the intervention and properly monitor providers. Citizen may also choose not to pay attention (Sims, 1998, 2003, 2006), often describe as rational inattention. As a consequence, introducing new information does not always lead to new beliefs or changes in behaviour.</p>	<p>If citizens' opportunity cost of paying attention to the information is high or they lack of attention span, their probability of participation will decrease<sup>iv</sup>.</p>
<p>High Opportunity Cost of Participation</p>	<p>Citizens, and particularly the poor, simply may not have the time to get informed or give feedback on service delivery because of more pressing priorities such as securing food and meeting other basic needs (Banerjee and Mullainathan, 2008).</p>	<p>If opportunity cost of participation is high, probability of participation will be lower.</p>
<p>Collective Action Failure</p>	<p>Scholars have emphasised the collective action problems that can arise in the presence of a non-</p>	<p>If citizens expect other citizens would free-ride on</p>

	excludable local public good (Olson, 1971), such as community monitoring. If community members believe fellow citizens will contribute to monitor the project, they may decide not to participate.	their efforts to monitor the project, the probability and intensity of participation will be lower.
Citizens' Beliefs can prevent participation	Citizens may refuse to take advantage of the opportunity to influence politicians and providers if they believe the chances of success are low. These beliefs can become a self-fulfilling prophecy where citizens refuse to participate and as a consequence providers have fewer incentives to improve performance (Molina, 2013a).	Citizens who perceived politicians and/or providers are responsive to them have higher probability of participation in community monitoring activities.
Elite Capture	Community monitoring may also be prone to be captured by local elites (Bardhan, 2002; Bardhan and Mookherjee, 2006; Olken, 2007). The rationale is that when decision making is part of the CMI, the elite would want to take advantage by capturing the monitoring process and appropriate the resources associated with the program.	If the CMI is captured by local elites, the participation will be limited to its supporters, which may affect the effectiveness of the program. It is an empirical question whether the elite capture could improve or worsen outcomes. See Atlas <i>et al.</i> (2013) for an example of different types of elites

## Politicians and providers' accountability

Under this heading we present potential reasons for a lack of responsiveness on the part of the politicians and providers. The literature cites many reasons why politicians and providers may not be accountable to their citizens (building block 4 and 6).

Below we identify three potential bottlenecks.

Table 3: Bottlenecks causing a lack of responsiveness from politicians and service providers

Bottleneck	Description	Empirical Implications
Unresponsive Politicians	Even in well-functioning democracies, citizens in a given community may not be pivotal for politician's electoral strategy (Downs, 1957; Hotelling, 1929; Persson and Tabellini, 2002). This means that citizens' support is not needed for politicians to win elections and/or stay in power. Additionally, especially in developing countries, often the political system does not work properly and institutions do not help translate the preference of the people into policy (Boix <i>et al.</i> , 2003; Acemoglu and Robinson, 2008). Keefer and Khemani (2004, 2005) argue that public service providers have weak incentives to improve performance quality because their jobs are protected by political agents – politicians have stronger incentives to provide secure public-sector jobs as teachers, health workers, and local bureaucrats, than to pressure these job-holders to improve service delivery.	If the community is not needed for the politicians to stay in power, we should find that politicians' performance does not increase as a result of the CMI, irrespective of what happens with citizen engagement in monitoring activities.
Unresponsive Providers	The literature on providers' motivations to deliver services no longer assumed them to be either public spirited altruists (knights) or passive	In communities where providers are not responsive to politicians, CMIs will only be



---

recipients of state largesse (pawns). Instead, they	effective if it changes providers'
are often considered to be in one way or another	behaviour.
self-interested (knaves) (Le Grand, 2003).	If communities can impose a
Communities in developing countries often have low	credible threat of informal social
state capacity, which limits the ability of	sanctions to unresponsive
governments to monitor self-interested providers	providers, the probability of a
(Besley and Persson, 2011). If this is the case,	change in behaviour from
putting pressure on the government will be	providers will be higher,
ineffective and only competition or informal	regardless of whether they are
sanctions from the community may have an effect	responsive to politicians.
on providers' performance.	If communities can choose
	providers, competition among
	them will foster better
	performance <sup>v</sup> .

---

### 3. Approach

We followed Campbell and Cochrane Collaboration guidelines for systematic reviews and meta-analysis (Shadish and Myers, 2004; Hammerstrøm et al., 2010; Booth, 2011; Higgins & Green, 2011; Campbell Collaboration, 2015) and were also informed by approaches developed by 3ie (White, 2009; Snilstveit, 2012; Waddington *et al.*, 2012).

We aimed to answer two specific review questions:<sup>vi</sup>

- 1) What are the effects of CMIs that use encouragement of the community to monitor service delivery, on service delivery and corruption in L&MICs?
- 2) What are the mechanisms through which CMIs effect a change in service delivery and corruption outcomes?

We searched for eligible studies of CMI in any L&MIC country defined according to the World Bank definition at the time the intervention being studied was carried out. We searched for relevant studies across a broad range of online databases, websites and knowledge repositories which allowed the identification of both peer reviewed and grey literature. Keywords for searching were translated into Spanish, French, and Portuguese and relevant non-English language literature was included. We also conducted reference snowballing and contacted experts and practitioners to identify additional studies. We used Endnote software to manage citations, abstracts, and documents. First stage results were screened against the inclusion criteria by two independent reviewer with additional supervision by a third.

Included CMIs needed to have a clear objective of reducing corruption and/or improving service delivery, use encouragement of the community to monitor service delivery as a key intervention instrument and fall into one of the four intervention categories defined above. Eligible comparisons could be passive, such as no formal process of monitoring (e.g. Björkman and Svensson, 2009), or active, for example CMI with no encouragement to participate (e.g. Olken, 2007).

We included quantitative studies with either experimental or quasi experimental design to address review question 1. In addition, both quantitative and qualitative studies were eligible for inclusion to address review question 2, provided the evidence was taken from a study eligible for inclusion for question 1 or a 'sibling study' (that is, a study on the same intervention as those included for question 1).

Corruption and service delivery were the primary outcomes of the review. We defined corruption as dishonest or fraudulent conduct by those in power. A big issue in the literature is the difficulty in measuring corruption accurately (Pande & Olken, 2011). As a consequence, we included corruption estimates from both the forensic

economic literature (Zitzewitz, 2012) as well as measures based on perceptions of corruption. We also defined service delivery outcomes as access to and quality of the service. For example, if the goal of the CMI was to facilitate household access to clean water, the percentage of access to clean water and water quality would be the outcome of interest. If the goal was to monitor school performance, children's tests scores would be the desired outcome. We also collected data on intermediate outcomes measuring changes in behaviour, such as whether participants contribute to monitoring of the service or project and the behaviour and performance of providers and politicians

Two reviewers independently coded and extracted data on study details, design and relevant results from the included studies.<sup>vii</sup> Studies were critically appraised for potential bias across using a tool provided by 3ie (available in Waddington et al., 2014). 'Low risk of bias' studies were those in which clear measurement of and control for confounding was made, where intervention and comparison groups were described adequately and risks of spillovers or contamination were small, and where reporting biases and other sources of bias were unlikely. 'Medium risk of bias' studies had suspected threats to validity of the attribution methodology, or possible risks of spillovers or contamination, arising from inadequate description of intervention or comparison groups or suspected reporting biases. We did not classify any included studies as being of 'high risk of bias'.

We used a combination of statistical meta-analysis and narrative synthesis along the causal chain to pool the evidence. To prepare the data for meta-analysis we calculated standardized mean differences (SMD) and average standardized effects (ASE)<sup>viii</sup> and 95 percent confidence intervals (95% CIs) for continuous outcome variables, and risk ratios and risk differences and 95% CIs for dichotomous outcome

variables. All effect sizes were computed as continuous outcomes, excepting those from Pandey *et al.* (2007), which were computed as dichotomous outcomes.

We then synthesized results using statistical meta-analysis. Where possible we extracted data on intermediate outcomes such as citizen participation and public officials and service providers' responsiveness.

Where studies reported the effect of more than one intervention, we chose the interventions that could be classified as one of our four categories. In cases where more than one intervention was relevant, we pooled their effects before integrating them into meta-analysis, taking into account the correlation of the treatment and control groups between study arms to address possible dependency<sup>ix</sup>. This is the case of Afridi and Iversen (2013), Banerjee *et al.* (2010), Barr *et al.* (2012), Olken (2007) and Pradhan *et al.* (2014). Where we identified different measures for the same outcome, we followed a similar procedure. We computed a synthetic effect size, defined as the mean effect size in that study, with a variance that takes account of the correlation among the different outcomes (Borenstein *et al.*, 2009). Finally, when we found different follow up periods for comparable interventions, we compared them considering similar horizon time. This is the case for Björkman and Svensson (2009) and Björkman, de Walque and Svensson (2013), who report both the short and the medium term impact of an intervention in Uganda, and the short term impact of another intervention in the same place. In these cases, we only run meta-analysis for short term effects.

Further details about the methods used to undertake the review are provided in the published protocol (Molina *et al.*, 2014) and the full technical report (Molina *et al.*, 2016).

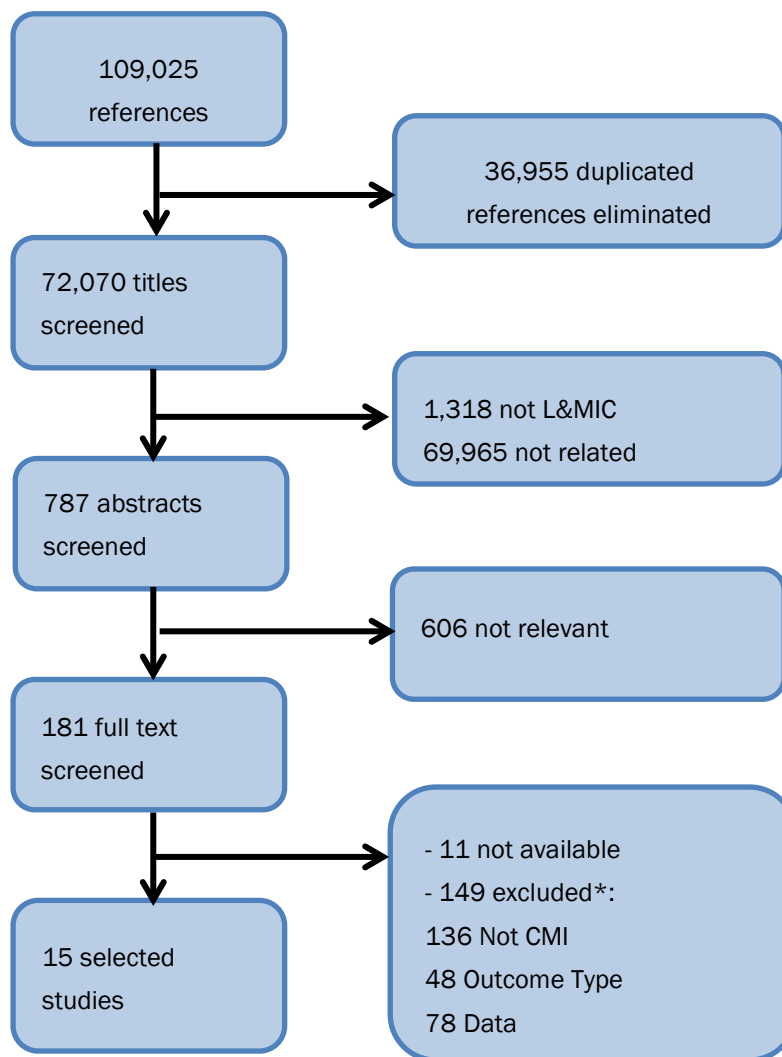
#### **4. Included study characteristics**

The search process (Figure 2) returned 109,017 references, of which 36,955 were eliminated as duplicates and a further 71,283 were excluded at the title screening stage. The remaining 787 papers were included for abstract screening and 181 studies were included for full text screening. Fifteen studies met the inclusion criteria for addressing review question 1 and six 'sibling studies' were also included to address question 2.

Descriptive information on included studies is provided in the Appendix table. Studies evaluated the effects of 23 different CMIs in the areas of information campaigns (10), scorecards (3), social audits (5), and combined information campaigns and scorecards (2). Included studies were conducted in Africa (6), Asia (7) and Latin America (2). Uganda and India had the largest presence of CMI impact evaluations, with four studies conducted in each country. We also identified two studies from Indonesia and one each from Benin, Liberia, Colombia, Pakistan and Mexico. Most studies focused on interventions in the education sector (9), followed by health (3), infrastructure (2) and employment promotion (1).

Studies assessed the effects of CMIs on all primary outcomes. Eleven studies reported on improvement in prevalence condition, seven on access to service, three on perception of corruption, two on average waiting time to get the service and two on forensic economic estimates for corruption. Different outcome measures were used by sector. For example, in health, access was measured as utilization, coverage or immunization. Ten studies used randomized assignment and five used quasi-experimental methodologies such as instrumental variables and matching (Keefer and Khemani, 2011; Molina, 2013b).

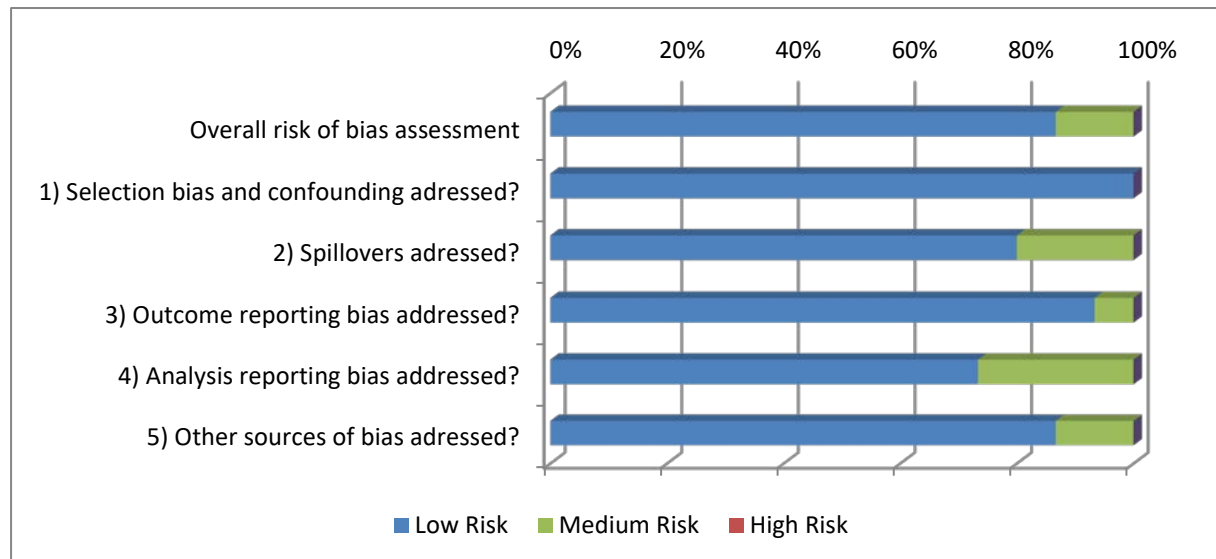
Figure 2: Search and selection process



Seven studies were categorised as of low risk of bias in attributing outcomes to the intervention, based on our five criteria of selection bias and confounding, spillovers, outcome reporting bias, analysis reporting bias, and other sources of bias (Andrabi *et al.*, 2013; Banerjee *et al.*, 2010; Barr *et al.*, 2012; Björkman & Svensson, 2009; Olken, 2007; Pandey *et al.*, 2007, and Pradhan *et al.*, 2014). The remaining eight studies were classified as of medium risk (Afridi & Iversen, 2013; Björkman *et al.*, 2013; Gertler *et al.*, 2008; Keefer & Khemani, 2011; Molina, 2013b; Pandey *et al.*, 2009; Pandey, Goyal & Sundararaman, 2009, Piper & Korda, 2010, and Reinikka & Svensson, 2011). Thus, the overall risk of bias assessment is predominantly low, with

13 out of 15 papers having this level of risk, followed by 2 papers with medium risk of bias (Figure 3).

*Figure 3: Summary of critical appraisal of studies of effects*



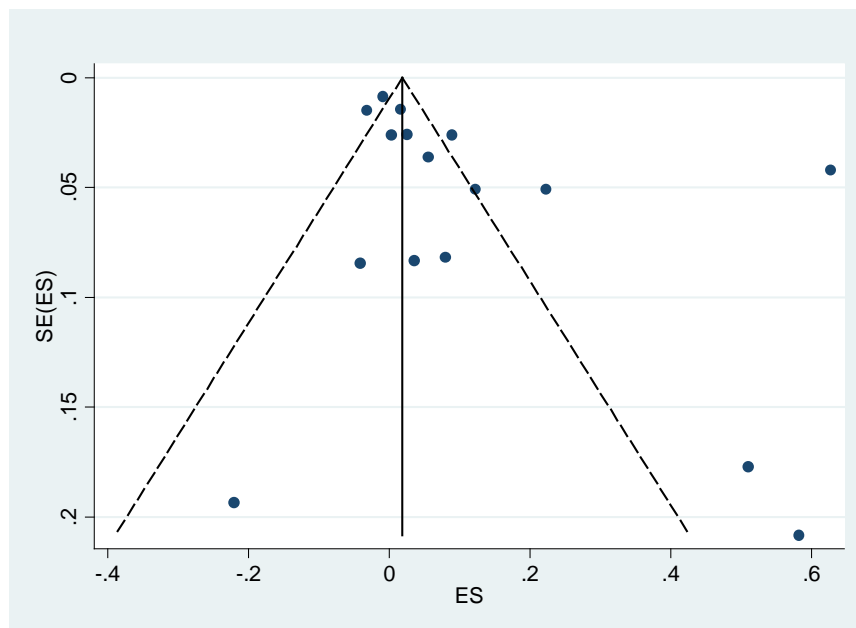
The majority of studies (11 out of 15) were adequately protected against performance bias as the units of treatment were located far from the control units. While in some cases the comparison group was selected from villages where the intervention was not carried out but were located near villages that had received the intervention<sup>x</sup>, and in other cases, the comparison group received a different treatment or the same intervention with a different degree of intensity (for example, Afridi and Iversen, 2013; Olken, 2007), the authors took that into consideration while designing the intervention and selecting cluster for their standard errors. We found just one case of potential outcome reporting bias, where the outcome reported was a new type of literacy test and the authors had not clearly justified the reason for using the measure over standard ones (Keefer and Khemani, 2011). With regards to analysis reporting bias, in most of the included studies different measures for the same outcome are reported or different specification and estimation methods are applied. There was no evidence

in outcomes were selectively reported and the authors used “common” methods of estimation.

We also assessed publication bias in the body of evidence, using tests for small study effects. We pooled all SMD effect sizes (reversing the sign when needed, so all measured positive effects in the same direction) and present the results in Figure 4.

The asymmetry in the plot, together with the result of the Egger (1997) Test (p-value = 0.058) suggests possible evidence of underreporting of small sample studies, which is often considered as evidence of publication bias. However, the low power of the tests prevents us from drawing strong conclusions.

*Figure 4: Funnel plot showing pseudo-95% confidence limits*



## **5. Impacts of CMLs (review question 1)**

In this section, we present the results of synthesis of effects to assess whether the included interventions reduced corruption and improved service delivery outcomes (review question 1). The included studies use a range of different measures to assess primary outcomes and it would not be sensible to pool conceptually different



outcomes in our meta-analysis. To avoid this problem we grouped studies only when the outcome variables represent a similar broad outcome construct and the intervention is implemented in the same sector. In the case of service delivery, we differentiated access from quality, and performed separate analysis by sector and outcome.

### **Corruption outcomes**

We identified few studies assessing the effect of CMI on corruption outcomes, using either forensic estimates (two studies, two interventions) or perception measures (three studies, four interventions). This limits our ability to generalize these results.

Included studies on the effects of CMI on corruption outcomes were implemented in infrastructure, education and employment assistance programs. Two studies used forensic economic estimates to collect outcomes data, and three studies (comprising four interventions) measured perception of corruption.

We looked for different measures of corruption in the papers considered, with the aim of extracting measures based on the application of economics to the detection and quantification of behavior (Zitzewitz, 2012), in this case, corruption. With this purpose, we extracted all measures that we could identify for each intervention.

Table 2 lists the measures reported in each case.<sup>xi</sup> We identified two studies reporting forensic measures of corruption (Olken, 2007 and Reinikka and Svensson, 2011). Olken (2007) evaluates the impact of increasing citizen participation in social audits in Indonesia on corruption, with two different treatment arms testing different variations of social audits. Villagers were invited to participate in social audits ("accountability meetings") in both treatment arms, but in one group the invitation was accompanied by an anonymous comment form, which could be submitted in a sealed box. The results of this exercise were summarized in the accountability meetings<sup>xii</sup>.

The study reports a forensic measure of corruption which is the percentage of missing funds in roads and ancillary projects. The effect of social audits only was SMD 0.08 (95% CI [-0.10, 0.26]) while the effect of social audits with anonymous comment form was also 0.08 (95% CI [-0.10, 0.25]). The combined effect for both treatment arms was SMD 0.08 (95% CI [-0.08 - 0.24]).<sup>xiii</sup>

Reinikka and Svensson (2011) evaluate the effect of systematic publication of monthly financial transfers to schools in Uganda. They found that a school close to a newspaper outlet suffers less from the capture of funds as compared to a school away from a newspaper outlet (Reinikka and Svensson, 2011).<sup>xiv</sup> The SMD shows that schools where the intervention took place had 22 percent less corruption than the others.

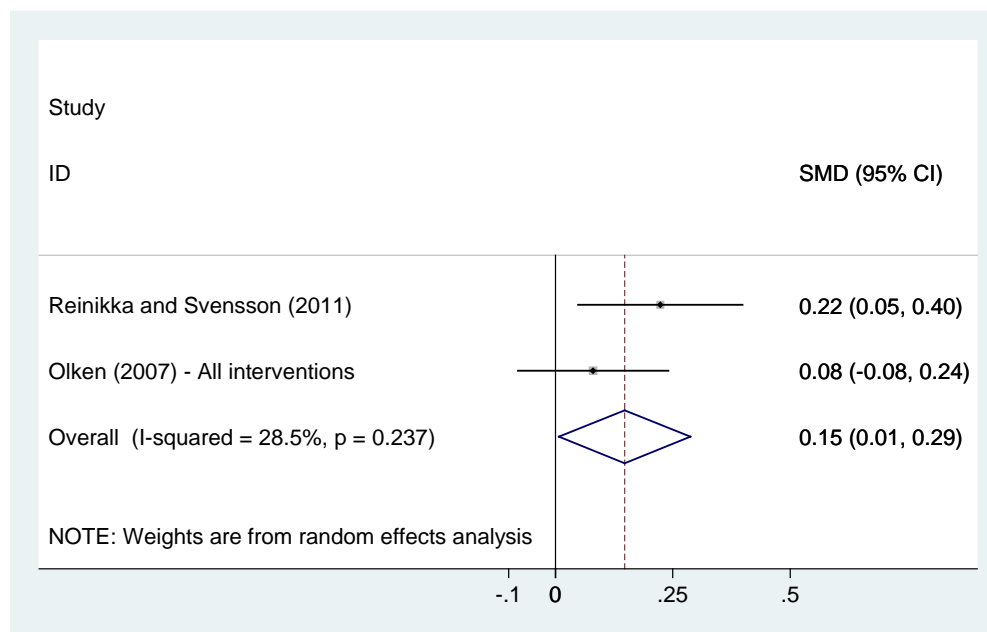
*Table 4: Forensic economic estimates of corruption*

<i>Study</i>	<i>Variable definition</i>	<i>CMI Type</i>	<i>Effect Size</i>	<i>95% Confidence Interval</i>		<i>ES Type</i>
Olken (2007) - Invitations	Percent missing funds*	Social Audit	0.08	-0.10	0.26	SMD
Olken (2007) - Invitations + comments	Percent missing funds*	Social Audit	0.08	-0.10	0.25	SMD
Olken (2007) - All interventions			0.08	-0.08	0.24	SMD
Reinikka and Svensson (2011)	Share of funding reaching school	IC	0.22	0.05	0.40	SMD
<b>Meta-analysis</b>			<b>0.15</b>	<b>0.01</b>	<b>0.29</b>	<b>SMD</b>

*Note: \* calculated as log reported value - log actual value (major items in roads and ancillary projects).*

Meta-analysis suggests that overall effect of these interventions is positive, improving corruption outcomes by 0.15 SD, as it shown in Figure 5. Since Olken (2007) finds no statistically significant effects, this result is probably driven by Reinikka and Svensson (2011), who did find a positive and statistically significant effect.

Figure 5: Corruption – forensic economic estimates



It is difficult to detect and measure corruption objectively and because of that we included more subjective perception measures. Table 5 lists the outcome measures reported in the three studies (four interventions) that we have included in this category. We were able to compute RD for the first two studies and SMD for the third one, so we analyze them by separate.

Table 5: Perception measures of corruption

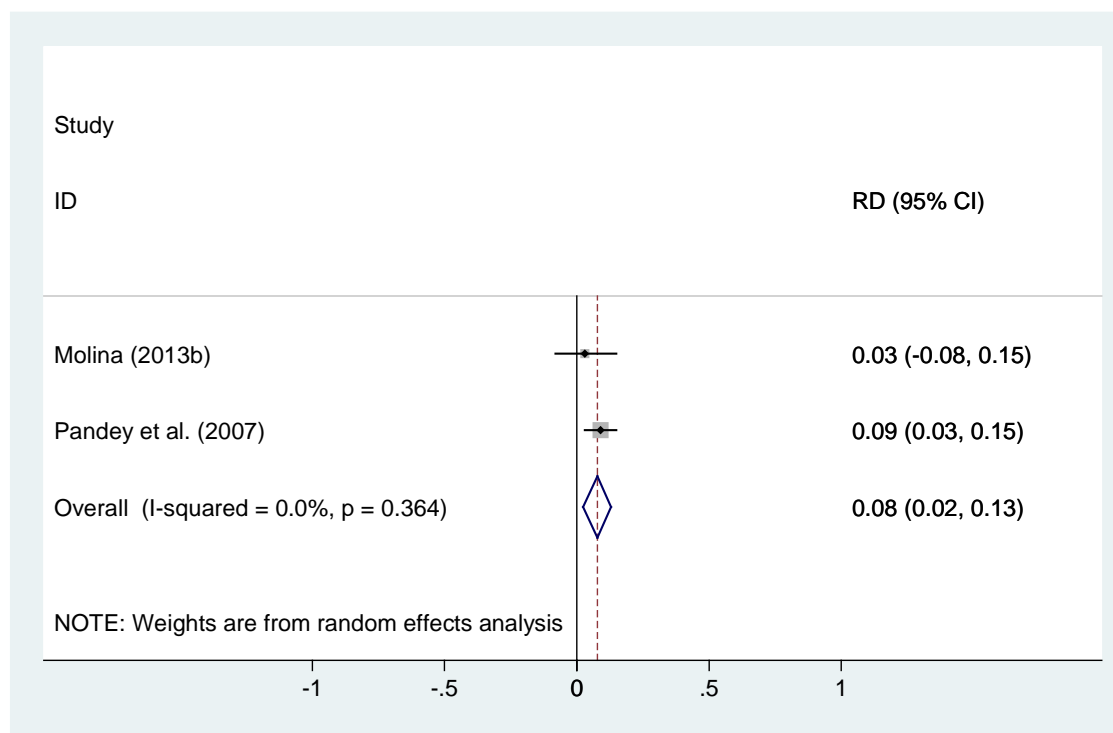
Study	Variable definition	CMI Type	Effect Size	95% Confidence Interval		ES Type
Molina (2013b)	Adequacy in the Administration of Resources	Social Audit	0.03	-0.08	0.15	RD
Pandey <i>et al.</i> (2007)	Percentage of household reporting that development work has been performed in the village	IC	0.09	0.03	0.15	RD
<b>Meta-analysis</b>			<b>0.08</b>	<b>0.02</b>	<b>0.13</b>	<b>RD</b>

Afridi and Iversen (2013) - Second audit	Total number of irregularities*	Social Audit	-0.22	-0.39	-0.05	SMD
Afridi and Iversen (2013) - Third audit	Total number of irregularities*	Social Audit	-0.23	-0.43	-0.04	SMD
<i>Afridi and Iversen (2013) - All interventions</i>			<i>-0.23</i>	<i>-0.38</i>	<i>-0.07</i>	<i>SMD</i>

*Note: \* reversed sign for this variable.*

We identified two studies assessing the effect of CMI on corruption perception measures for which comparable effect size were available. Molina (2013b) found evidence that a social audit improved the perception of the administration of resources in Colombia, while Pandey *et al.* (2007) suggest that an information campaign carried out in India increased the probability of households reporting that development work took place in the villages, which can be interpreted as a reduction in corruption. The overall effect is RD 0.08 (95% CI [0.02, 0.13]), as can be seen from the forest plot presented in Figure 6. The confidence intervals are overlapping and the test of homogeneity does not suggest between study variability. Both studies report a reduction in the perception of corruption among beneficiary households.

Figure 6: Corruption – perception measures



Afridi and Iversen (2013) assess the effect of social audits in India. They estimate the effect on reported irregularities in Gram Panchayats with one audit as compared to those with two and three audits respectively. The effect after two audits was to reduce irregularities by 0.22 SD (95% CI [-0.39, -0.05]) and after three audits by 0.23 SD (95% CI [-0.43, -0.04]), suggesting increased corruption in *Gram Panchayats* after these interventions, and a worsening effect with two or more social audits, as compared to those with one audit only. The authors explain that this is likely due to increasing complaints over time as participants become more aware of their entitlements, or have greater confidence in the integrity of the audit process, or if the quality of the social audit improves to increase the number of irregularities identified. (Afridi and Iversen, 2013).

### Access to services

A number of different outcome measures were included as proxies for access to services. In this section we begin with health services and present the results for utilization, immunization and other measures of access, followed by results for enrolment and dropout rates in the education sector.

### *Health*

We identified two studies that assessed at the impact of CMIs on utilization.

Björkman and Svensson (2009) evaluate the same intervention, a combination of a scorecard with an information campaign both in the short and in the medium term<sup>xv</sup>, using the same group of 50 facilities/communities that were randomly assigned to treatment and control group.

In addition Björkman, de Walque and Svensson (2013) assess a new intervention, an information campaign with new treatment and control groups in which 25 new facilities were randomly assigned to a treatment group (13 units) and control group (12 units). This intervention differs from the previous one since it does not include a scorecard with relevant information about the health service provision.

In these studies, authors report an ‘average standardized treatment effect’ following Kling *et al.* (2004)’s methodology, that can be combined into one meta-analysis given their homogeneity and given that they are comparable as they all imply better access to health services. Table 6 presents the effect size for each intervention regarding utilization of health services<sup>xvi</sup>.

*Table 6: Utilization*

<i>Study</i>	<i>Variable definition</i>	<i>CMI Type</i>	<i>Effect Size</i>	<i>95% Confidence Interval</i>	<i>ES Type</i>
--------------	----------------------------	-----------------	--------------------	--	----------------

Björkman and Svensson (2009) - Short Term	Utilization/coverage*	Scorecard + IC	2.13	0.79	3.47	ASE
Björkman, de Walque and Svensson (2013) - Short Term	Utilization/coverage*	IC	0.04	-0.41	0.49	ASE
<b>Meta-analysis</b>			<b>0.99</b>	<b>-1.05</b>	<b>3.02</b>	<b>ASE</b>
Björkman and Svensson (2009) - Medium Term	Utilization/coverage*	Scorecard + IC	0.34	0.12	0.55	ASE

*Note: \* weighted mean of average number of patients visiting the facility per month for out-patient care, average number of deliveries at the facility per month, share of visits to the project facility of all health visits averaged over catchment area, and share of visits to traditional healers averaged over catchment area.*

Looking at the short run, both interventions show an increase in the access to health services, although for the second one the result is statistically no significant. This suggests that effects are stronger when the information campaign is coupled with a scorecard, which is consistent with the authors' findings, who hint that without information, the process of stimulating participation and engagement had little impact on health workers' performance or the quality of health care (Björkman, de Walque and Svensson, 2013). The overall effect is positive but not statistically significant, and there is a large amount of between study variability ( $I^2 = 88.0\%$ ,  $p=0.004$ ).

Looking at the medium term, the information campaign combined with the scorecard has a positive and statistically significant effect on service access. Regarding immunization outcomes, Table 7 reports the short run effects found by Björkman, de Walque and Svensson (2013) and Pandey *et al.* (2007). It also reports the medium term effects for the intervention assessed in Björkman and Svensson (2009) which is not incorporated into meta-analysis given the different time horizons.

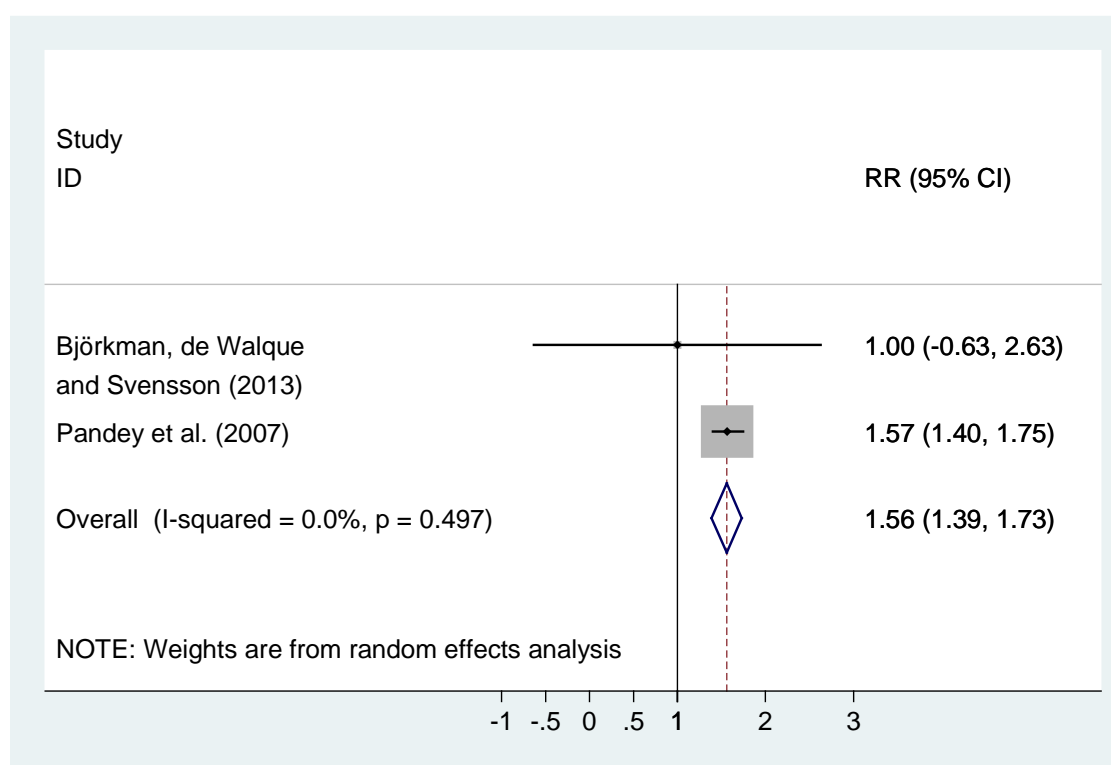
Table 7: Immunization

Study	Variable definition	CMI Type	Effect Size	95% Confidence Interval		ES Type
Björkman, de Walque and Svensson (2013)	Immunization*	IC	1.00	-0.63	2.63	RR
Pandey <i>et al.</i> (2007)	Vaccinations received by infants	IC	1.57	1.40	1.75	RR
<b>Meta-analysis</b>			<b>1.56</b>	<b>1.39</b>	<b>1.73</b>	<b>RR</b>
Björkman and Svensson (2009) - Medium Term	Immunization*	IC	1.04	-0.52	2.61	RR

Note: \* mean of newborns, children less than 1-year, 1-year olds, 2-year olds, 3-year olds and 4-year olds, whether the child has received at least one dose of measles, DPT, BCG, and Polio.

Björkman, de Walque and Svensson (2013) assess the impact of the CMI on immunization for different age groups, while Pandey *et al.* (2007) compute the percentage of households where infants have received vaccinations. The overall effect is an increase in risk ratio of 1.56 (95% CI [1.39, 1.73]), implying that the effect of the CMI improved access to services by 56 percent (Figure 7). The medium term impact of the intervention reported by Björkman and Svensson (2009) is 1.04 SD but statistically non-significant.<sup>xvii</sup>

Figure 7: Immunization





Pandey *et al.* (2007) also report on different measures of access to health services, specifically the percentage of households getting health services such as visits by nurse midwives, prenatal examinations, tetanus vaccinations, and prenatal supplements received by pregnant women. We computed risk ratios for these outcomes, and the results are reported in Table 8. All risk ratios are above unity, with an overall effect of 1.43 (95% CI [1.29, 1.58]) implying that the intervention improved access to services in 43 percent.

*Table 8: Other access to service outcomes*

Study	Variable definition	CMI Type	Effect Size	95% Confidence Interval		ES Type
Pandey <i>et al.</i> (2007)	Visits by nurse midwife	IC	1.03	0.94	1.14	RR
	Prenatal examinations		1.63	1.45	1.83	RR
	Tetanus vaccinations		1.57	1.39	1.77	RR
	Prenatal supplements received by pregnant women		1.45	1.29	1.64	RR
	Vaccinations received by infants		1.57	1.40	1.75	RR
<b>Meta-analysis</b>			<b>1.43</b>	<b>1.29</b>	<b>1.58</b>	<b>RR</b>

## *Education*

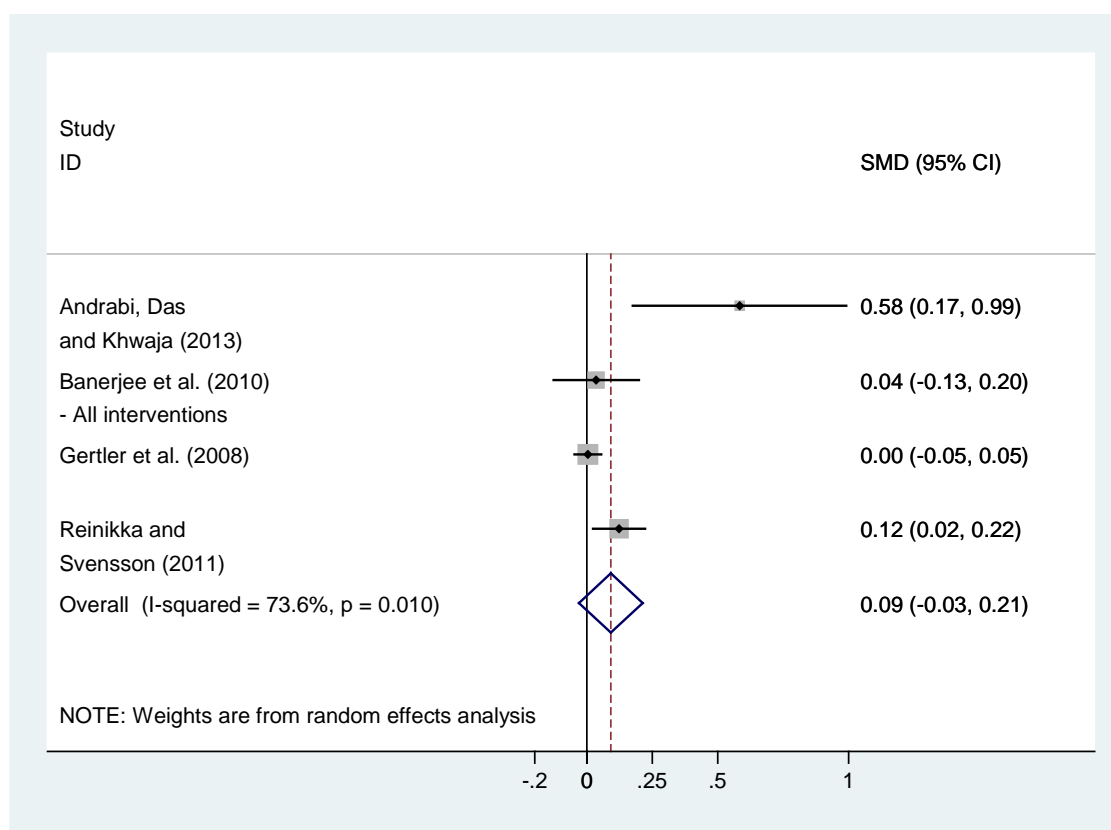
We identified four studies evaluating effects on enrolment in six different treatment arms. Table 9 presents the effect sizes from all treatment arms. Before combining the studies into a meta-analysis, we created synthetic effect sizes for the study with multiple treatment arms (Banerjee *et al.*, 2010) to avoid combining effects based on dependent samples. Figure 8 presents the forest plot for the meta-analysis of enrolment rates. The overall average effect of CMI on enrolment is 0.09 SD (95% CI [-0.03, 0.21]). However, it can be noted that this result is driven by the inclusion of one

study for which the SMD is substantially higher than the others (Andrabi, Das and Khwaja, 2013). Also, there is significant between study heterogeneity ( $I^2 = 73.6\%$ ,  $p=0.010$ ). Therefore, we performed sensitivity meta-analysis excluding this study, which yielded an overall effect of 0.05 SD (95% CI [-0.03, 0.13]).

*Table 9: Enrollment*

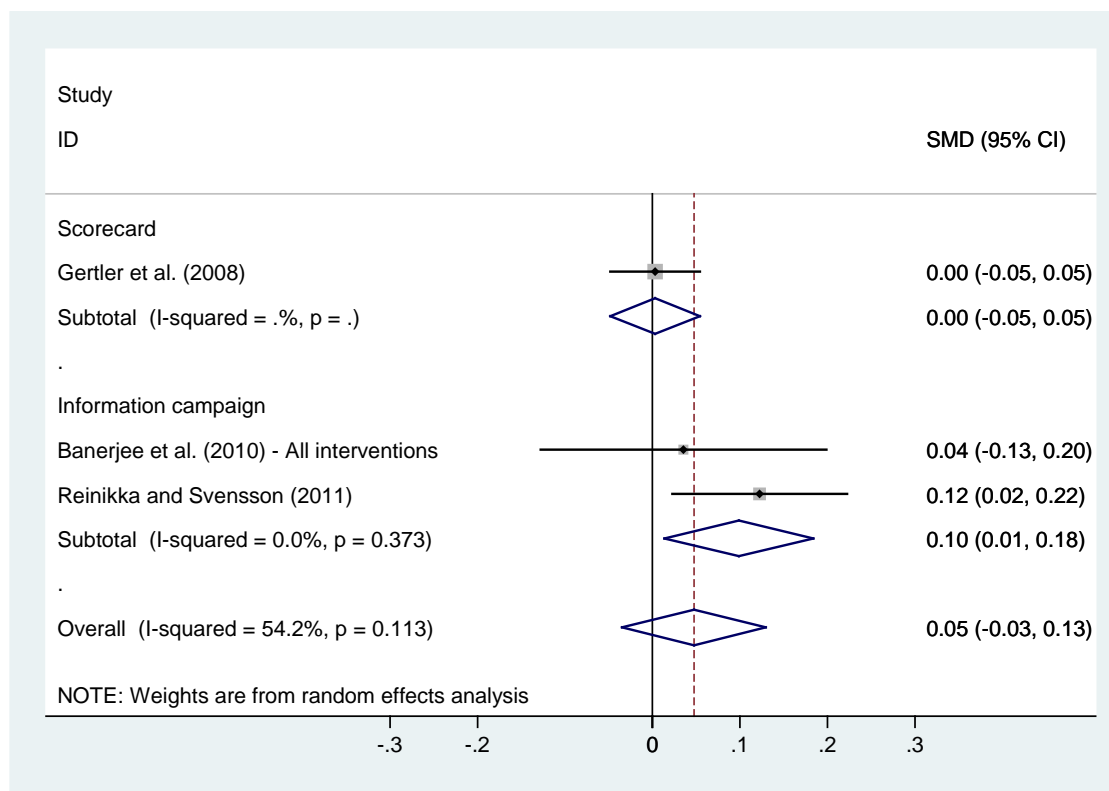
<i>Study</i>	<i>Variable definition</i>	<i>CMI Type</i>	<i>Effect Size</i>	<i>95% Confidence Interval</i>		<i>ES Type</i>
Andrabi, Das and Khwaja (2013)	Enrolment	Scorecard	0.58	0.17	0.99	SMD
Banerjee <i>et al.</i> (2010) - Mobilization	Enrolment	IC	0.059	-0.138	0.257	SMD
Banerjee <i>et al.</i> (2010) - Mobilization + information	Enrolment	IC	0.05	-0.14	0.25	SMD
Banerjee <i>et al.</i> (2010) - Mobilization + information + "Read India"	Enrolment	IC	-0.008	-0.199	0.183	SMD
<i>Banerjee et al. (2010) - all interventions</i>			<i>0.04</i>	<i>-0.13</i>	<i>0.20</i>	<i>SMD</i>
Gertler <i>et al.</i> (2008)	Enrolment	Scorecard	0.003	-0.048	0.054	SMD
Reinikka and Svensson (2011)	Enrolment	IC	0.12	0.02	0.22	SMD
<b><i>Meta-analysis</i></b>			<b><i>0.09</i></b>	<b><i>-0.03</i></b>	<b><i>0.21</i></b>	<b><i>SMD</i></b>

Figure 8: Enrollment



When considering only scorecards (Gertler *et al.*, 2008), the overall effect is zero (SMD=0.003; 95% CI [-0.05, 0.05]). On the other hand, information campaigns show an overall effect SMD 0.10, 95% CI [0.01, 0.18], suggesting that these interventions have increased enrolment rates in 10 percent (*Figure 9*).

Figure 9: Enrollment – sensitivity analysis (outlier excluded)



We identified four studies measuring dropouts at schools in seven treatment arms (Table 10). For some interventions, the results suggest an increase in children out of school in the villages receiving CMI compared to those that did not receive the program. Considering the study of Banerjee *et al.* (2010), the effects range from 0.02 SD improvement (95% CI [-0.01, 0.06]) for the treatment arm with mobilization and information, to 0.05 SD (95% [0.011, 0.081]) for the treatment arm with mobilization and information in addition to “Read India” – reading camps held by trained volunteers – with a combined effect of 0.03 SD (95% CI [0.003, 0.06]). The authors argue that this result is due to “*children dropping out of private or NGO schools (results omitted to save space). It may be that parents consider the reading classes to be an adequate alternative to a private school*”. The CMI also resulted in an increase in dropout rates in the cases of Andrabi, Das and Khwaja (2013) and for the Training intervention in Pradhan *et al.* (2014). On the other hand, the Linkage

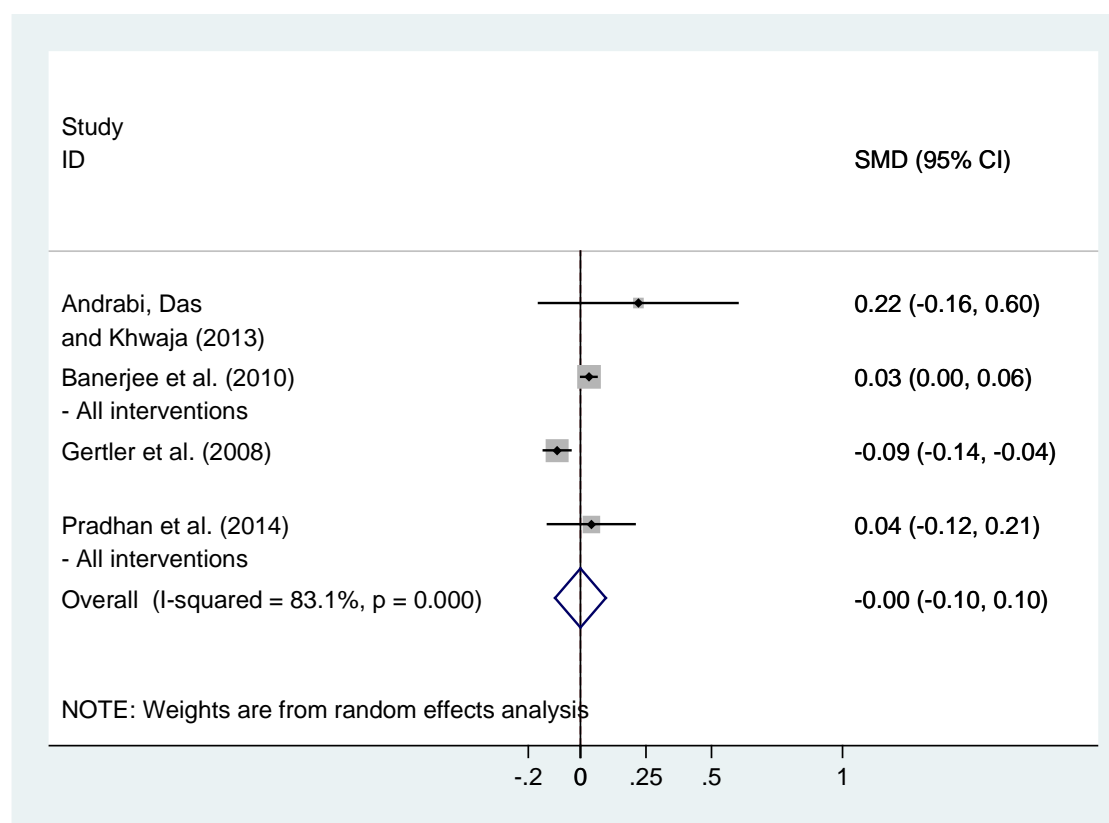
intervention in Pradhan *et al.* (2014) and the study of Gertler *et al.* (2008) find a reduction in dropout rates after interventions.

*Table 10: Dropouts*

<i>Study</i>	<i>Variable definition</i>	<i>CMI Type</i>	<i>Effect Size</i>	<i>95% Confidence Interval</i>		<i>ES Type</i>
Andrabi, Das and Khwaja (2013)	Dropout rate	IC	0.220	-0.159	0.600	SMD
Banerjee <i>et al.</i> (2010) - Mobilization	Dropout rate	IC	0.028	-0.006	0.063	SMD
Banerjee <i>et al.</i> (2010) - Mobilization + information	Dropout rate	IC	0.02	-0.01	0.06	SMD
Banerjee <i>et al.</i> (2010) - Mobilization + information + "Read India"	Dropout rate	IC	0.046	0.011	0.081	SMD
<i>Banerjee et al. (2010) - all interventions</i>			<i>0.032</i>	<i>0.003</i>	<i>0.061</i>	<i>SMD</i>
Gertler <i>et al.</i> (2008)	Dropout rate	IC	-0.09	-0.14	-0.04	SMD
Pradhan <i>et al.</i> (2014) - Training	Dropout rate	IC	0.12	-0.08	0.31	SMD
Pradhan <i>et al.</i> (2014) - Linkage	Dropout rate	IC	-0.03	-0.23	0.16	SMD
<i>Pradhan et al. (2014) - all interventions</i>			<i>0.041</i>	<i>-0.124</i>	<i>0.207</i>	<i>SMD</i>
<b><i>Meta-analysis</i></b>			<b><i>0.00</i></b>	<b><i>-0.10</i></b>	<b><i>0.10</i></b>	<b><i>SMD</i></b>

Before performing the meta-analysis (Figure 10), we calculated a synthetic effect size for the two treatment arms included in Pradhan *et al.*'s (2014) study from Indonesia to avoid issues with dependent effect sizes in the meta-analysis. We did the same with the three interventions reported by Banerjee *et al.* (2010). Taking into account all the interventions, the overall effect is zero (SMD=0.00; 95% CI [-0.10, 0.10]).

Figure 10: Dropouts



## Quality of services

In this section, we present the analysis of effects on quality of services by sector and outcome, starting with health, and followed by education.

### Health

For health related outcomes, we consider measures of child death and anthropometric outcomes. We identified two studies with measurements of child mortality, Björkmann and Svensson (2009) and Björkman, de Walque and Svensson (2013). Table 11 shows that the short term evaluation for the two interventions had an overall effect of RR 0.76 (95% CI [0.42, 1.11]), suggesting that child death had been reduced by 24 percent after CMIs, however the effect is not statistically significant. Similar conclusions apply for the medium term effect of the information

campaign combined with a scorecard, were the effect is a reduction in 21 percent in mortality.

*Table 11: Child death*

<i>Study</i>	<i>Variable definition</i>	<i>CMI Type</i>	<i>Effect Size</i>	<i>95% Confidence Interval</i>		<i>ES Type</i>
Björkman and Svensson (2009) - Short Term	Death (under five mortality rate)	Scorecard + IC	0.65	0.42	1.02	RR
Björkman, de Walque and Svensson (2013) - Short Term	Death (infant mortality rate)	IC	1.05	0.61	1.81	RR
<b>Meta-analysis</b>			<b>0.76</b>	<b>0.42</b>	<b>1.11</b>	<b>RR</b>
Björkman and Svensson (2009) - Medium Term	Death (under five mortality rate)	Scorecard + IC	0.79	0.57	1.08	RR

We can also interpret an improvement in anthropometric measures as an improvement in the quality of health services provided. Table 12 reports the impact of the same two CMIs on weight-for-age scores.

*Table 12: Weight for age*

<i>Study</i>	<i>Variable definition</i>	<i>CMI Type</i>	<i>Effect Size</i>	<i>95% Confidence Interval</i>		<i>ES Type</i>
Björkman and Svensson (2009) - Short Term	Weight for age (children 0-18 months)	Scorecard + IC	1.20	1.00	1.43	RR
Björkman, de Walque and Svensson (2013) - Short Term	Weight for age (children 0-11 months)	IC	1.22	0.92	1.60	RR
<b>Meta-analysis</b>			<b>1.20</b>	<b>1.02</b>	<b>1.38</b>	<b>RR</b>
Björkman and Svensson (2009) - Medium Term	Weight for age (children 0-18 months)	Scorecard + IC	1.29	1.01	1.64	RR

In the short term, CMIs have increased weight for age scores by 20 percent, suggesting that quality of health services has improved. The positive impact seems to be stronger in the medium term, resulting in a 29 percent improvement.

Homogeneity test suggests no variability between studies ( $I^2 = 0.01\%$ ,  $p=0.928$ ).

In addition to these measures of health services' quality, these studies also report on another measure, namely average waiting time in medical facilities. The effects range from RR 0.91 (95% CI [0.81, 1.01]) to RR 1.10 (95% CI [0.81, 1.15]) and are displayed in Table 13.<sup>xviii</sup> Meta-analysis for the short term interventions suggests a negligible effect (1 percent reduction in waiting time). However, there is a large between study heterogeneity that might be driving these results ( $I^2 = 70.8\%$ ,  $p=0.064$ ).

*Table 13: Average waiting time to get the service*

<i>Study</i>	<i>Variable definition</i>	<i>CMI Type</i>	<i>Effect Size</i>	<i>95% Confidence Interval</i>		<i>ES Type</i>
Björkman and Svensson (2009) - Short Term	Waiting time in medical services*	Scorecard + IC	0.91	0.81	1.01	RR
Björkman, de Walque and Svensson (2013) - Short Term	Waiting time in medical services*	IC	1.10	0.81	1.15	RR
<b>Meta-analysis</b>			<b>0.99</b>	<b>0.80</b>	<b>1.17</b>	<b>RR</b>
Björkman and Svensson (2009) - Medium Term	Waiting time in medical services*	Scorecard + IC	1.06	0.95	1.19	RR

*Note: \* Difference between the time the user left the facility and the time the user arrived at the facility, subtracting the examination time.*

## *Education*



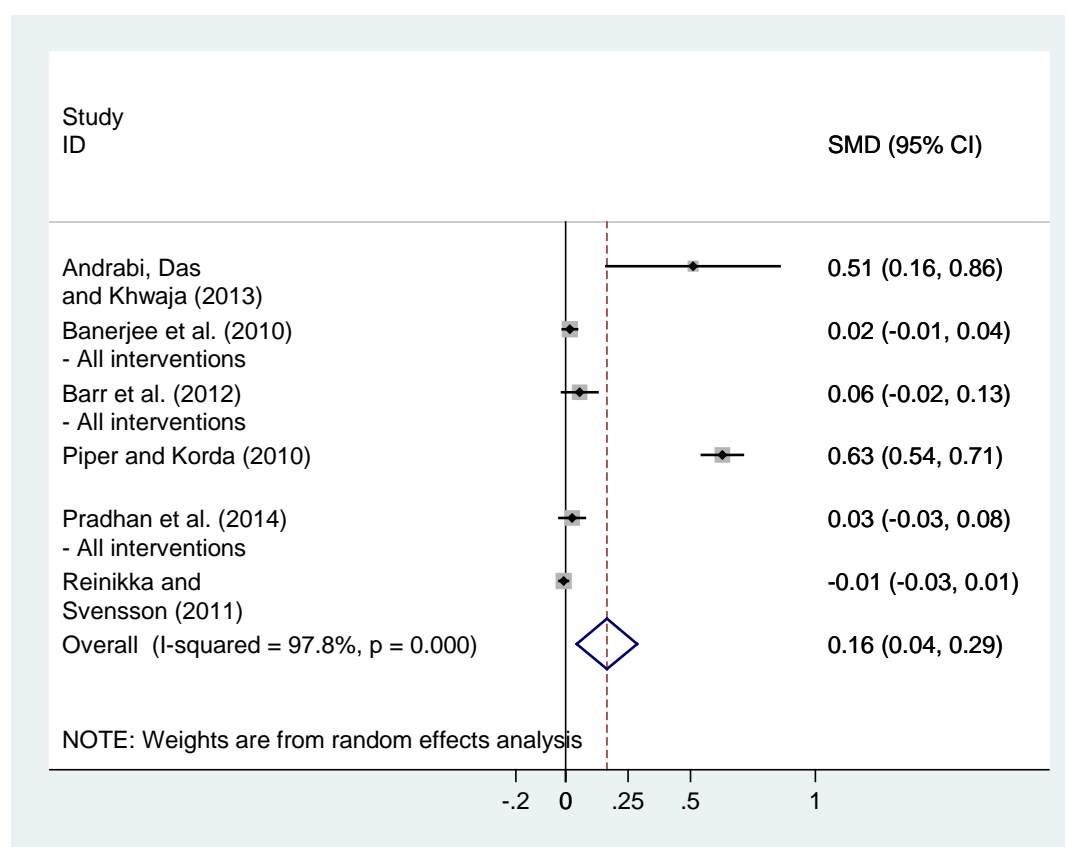
We included six studies assessing the effect of CMI on the quality of education as measured by test scores<sup>xix</sup>. As can be seen from Table 14, three of these studies include multiple treatment arms. We calculated synthetic effect sizes combining the different treatment arms before including these studies in the meta-analysis. The overall average effect of CMI on student outcomes across these six studies is an increase of 0.16 SD (95% CI [0.04, 0.29]).<sup>xx</sup>

*Table 14: Test Scores*

<i>Study</i>	<i>Variable definition</i>	<i>CMI Type</i>	<i>Effect Size</i>	<i>95% Confidence Interval</i>		<i>ES Type</i>
Andrabi, Das and Khwaja (2013)	test score	Scorecard	0.510	0.163	0.857	SMD
Banerjee <i>et al.</i> (2010) - Mobilization	test score	IC	0.01	-0.02	0.04	SMD
Banerjee <i>et al.</i> (2010) - Mobilization + information	test score	IC	0.010	-0.018	0.037	SMD
Banerjee <i>et al.</i> (2010) - Mobilization + information + "Read India"	test score	IC	0.03	0.00	0.05	SMD
<i>Banerjee et al. (2010) - All interventions</i>			0.02	-0.01	0.04	SMD
Barr <i>et al.</i> (2012) - Standard scorecard	test score	Scorecard	0.03	-0.05	0.11	SMD
Barr <i>et al.</i> (2012) - Participatory scorecard	test score	Scorecard (Participatory)	0.078	-0.002	0.158	SMD
<i>Barr et al. (2012) - All interventions</i>			0.056	-0.015	0.127	SMD
Piper and Korda (2010)	test score	IC	0.63	0.54	0.71	SMD
Pradhan <i>et al.</i> (2014) - Training	test score	IC	-0.02	-0.09	0.04	SMD
Pradhan <i>et al.</i> (2014) - Linkage	test score	IC	0.07	0.02	0.13	SMD
<i>Pradhan et al. (2014) - All interventions</i>			0.03	-0.03	0.08	SMD
Reinikka and Svensson (2011)	test score	IC	-0.01	-0.03	0.01	

The assessment of homogeneity suggests a large amount of variability between studies. This is further supported by the forest plot in Figure 11. The effect sizes range from SMD -0.01 (95% CI [-0.03, 0.01]) in Uganda (Reinikka and Svensson, 2011) to SMD 0.63 (95% CI [0.54, 0.71]) in Liberia (Piper and Korda, 2010). The confidence intervals of these two studies do not overlap.

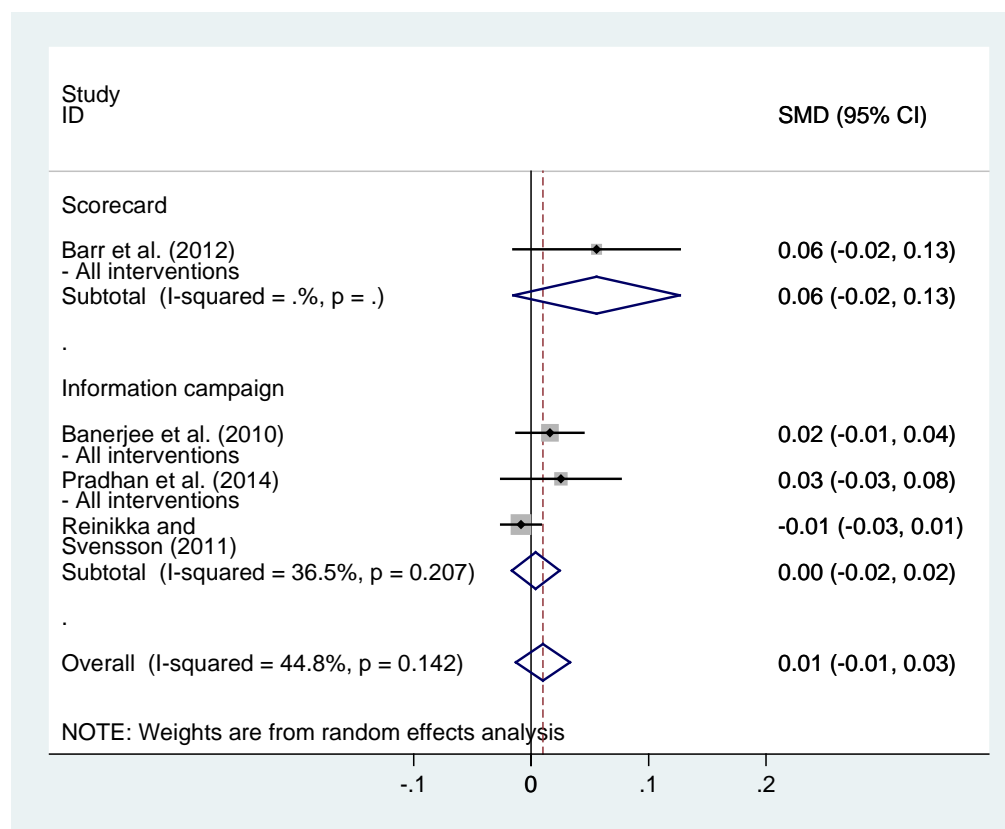
Figure 11: Test scores



When we exclude the outliers (Andrabi, Das and Khwaja, 2013 and Piper and Korda, 2010), the overall effect is SMD 0.01 (95% CI [-0.01, 0.03]).

We conducted sensitivity analysis by CMI type, excluding the possible outliers (Figure 12). The overall effect for information campaigns is zero (SMD=0.004; 95% CI [-0.017, 0.024]).

Figure 12: Test scores – sensitivity analysis (outliers excluded)



## 6. Analysis of causal mechanisms (review question 2)

In this section, we synthesise evidence on the causal pathways through which CMIs may have an effect on corruption and service delivery outcomes, drawing on data available in studies of effects and the 6 ‘sibling studies’ we were able to identify (Table A1). Sometimes causal pathways are explicit and other times finding pathways meant looking for implicit assumptions and arguments. We revised the causal chain to show the assumptions we were able to identify from this causal pathway analysis (Figure A1), and present findings here according to determinants of citizen participation and politician and provider accountability.

### *Citizen participation*

A potential concern with CMIs is that citizens may fail to participate in monitoring activities (building block 3). We identified potential bottlenecks that could prevent

citizens from participating in monitoring activities, which in turn reduces the potential impact of the program. In particular, if community monitoring activities are not carried out, or carried out by only a few citizens, the likelihood they uncover problems and put pressure on the government to provide accountability can be significantly reduced. We grouped evidence on citizens' participation in monitoring activities into five areas.

Firstly, **inadequate information on how to monitor the project** does not seem to be an important factor in explaining failures in participation in Colombia and India. The evidence also suggests that even when information is poor, providing that information is not sufficient to increase participation. In India, Banerjee *et al.* (2007) found that parents, teachers and VEC (local equivalent to parent-teacher association) members were not fully aware of how low students' performance was or the role of local committees, and how local participation might improve school performance. Many parents were not aware of the existence of the VEC (sometimes even when they were members of it), and VEC members were unaware of the key roles they were assigned in the educational system. The authors therefore examined different information interventions to mobilize parents to monitor schools, finding that providing information was not effective in encouraging participation. However, encouraging communities to provide volunteers who were subsequently trained by Pratham staff on techniques to teach children how to read and then running after-class reading classes for them, did improve reading levels by 3 to 8 per cent in treatment villages. Molina (2013b) also found that in some communities, citizens were not aware of the existence of the project they were supposed to monitor, and in some others, although they knew about the project, they lacked information on how to monitor it. This prevented citizens in those communities from taking an active role in social audit

community forums and community monitoring activities in general. However, this was not the only or even the most important barrier to participation in this study.

Second, evidence on whether **high opportunity costs of participation** could explain limited engagement varied. Andrabi, Das and Khwaja (2013) find that better educated parents participate more actively in monitoring activities in Pakistan, which might indicate that the opportunity cost of participation is lower for them. In contrast, Molina (2013b), studying the “Auditorias Visibles” social audit project in Colombia, was not able to find differences between participants and non-participants in employment status, income level or whether they work at home or not.

Third, **pessimistic beliefs about politician and provider responsiveness** were found in studies in Colombia and India. Molina (2013b) found that perceiving oneself as influential to local government is crucial for deciding whether to spend time in community monitoring activities. Banerjee *et al.* (2010) suggest citizens’ perceptions about providers’ responsiveness were important in determining engagement with social accountability interventions in India, since only the intervention that did not involve government action had significant effects on participation.<sup>xxi</sup> Also in India, Afridi and Iversen (2013) believed cases of maladministration and corruption in MGNREGA were underreported because of citizens’ pessimistic beliefs about the integrity of the audit process. Singh and Vutukuru (2009), analysing the initial stages of the same intervention, also describe a situation with little political enthusiasm during the pilot phase of the social audits, but with subsequent high level political support generating huge increases in the turnouts at the local social audit meetings.

Evidence suggested interactions between citizens and service providers could change citizens’ perceptions of low accountability. Barr *et al.* (2012) used the voluntary contributions mechanism game, a public goods game, played immediately

after training of school management committees, to show that willingness to contribute to public goods is significantly higher for those in participatory treatment arms. However, the authors were not able to discriminate between two potential mechanisms: that the participatory treatment affects preferences or beliefs about the willingness of providers to contribute to public goods.

Woodhouse (2005) analysed the early stages of the KDP program in Indonesia, finding that when villagers possessed information about their rights and, crucially, when the potential perpetrators of corruption knew that villagers had this information, it raised the perceived cost of corrupt behaviour and reduced the cost of fighting it. These pessimistic beliefs could also have been the cause of elite capture (Olken, 2007).

Finally, the effect of having **beliefs that other citizens would not participate** could not be verified. Björkman and Svensson (2010) suggest that citizen participation may be threatened by differences within the community, in particular because inequality and ethnic fractionalization may be barriers to collective action. However, the authors were unable to determine whether failures of collective action were due to lack of trust among community members, lack of trust of the community in the service providers and representatives, or both.

#### *Politician and provider accountability*

Providers and politicians need to benefit (by gaining popularity, increasing or maintaining salaries, and/or social recognition) in return for being responsive to citizens. If these assumptions are not met, the underlying program theory of the social accountability information breaks down, in particular by affecting citizens' decisions on whether to monitor government activity. According to our causal chain, even if participation is high, the CMI may be ineffective if politicians can ignore the

demand for accountability (e.g. the politicians does not need their support, clientelism, among others), or even when politicians are committed to improve service delivery, providers are non-responsive.

In this subsection we are interested in understanding why demand was lacking. Is it because (a) politicians would be unresponsive to demand, (b) providers would be unresponsive to increased pressure from politicians, or (c) citizens believe politicians/providers would be unresponsive. It is important to note the difference between (a) and (b) with (c). While in (a) and (b) politicians and providers respectively are not responsive, in alternative (c) they are responsive but citizens believe they are not. As a result they do not participate in monitoring and politicians and providers act as if there is no demand.

Molina (2013) suggests that when the community increases its demands by increasing citizen participation in the social audit, the politicians respond by performing better, as evaluated by the citizens, in this case in Colombia. Keefer and Khemani (2011) found that better learning outcomes were not due to better providers' performance, but due to changes in households' behaviour, as citizens decided to circumvent the existing institutions to monitor service providers and instead used private sector providers of education.

On the other hand, Björkman and Svensson (2009) found that scorecards were effective in improving waiting times, staff attendance and attitudes towards participants, because of specific innovations implemented to improve management as a result of the CMI. These included a suggestion box for complaints and recommendations, numbered waiting cards for patients, a poster informing about free health services, and a poster on patients' rights and obligations. Hence, according to

the authors, the program was successful due to the behaviour change among health facility staff.

Perhaps more revealing is the evidence from studies with multiple treatments, with and without facilitated (participatory) contact with providers. For example, Barr *et al.* (2012) found that only the facilitated contact treatment was effective in reducing teacher absence, compared to the standard treatment. Pradhan *et al.* (2014) also found evidence that facilitated contact between users and providers may enhance motivation for citizens to care about service outcomes and for providers to perform better, for some outcomes.

## **7. Discussion**

Table 15 summarizes the findings of the effects of CMIs. The results suggest positive effects on reducing corruption, whether measured using forensic economic estimates or through participant perception. In the case of health service delivery access, two interventions examined utilization, one finding a moderately large effect size in the medium term. For immunization, we identified a positive effect in the short term across two interventions, but not in the medium term.

Evidence suggests that CMIs by themselves do not improve access to services in education. Across six interventions looking at education enrolment, results indicate CMIs increase enrolment rates by a small but marginally statistically insignificant amount. Pooling across seven interventions also does not suggest CMIs reduce school dropouts. On the other hand, evidence from 10 studies suggests CMIs significantly improve test scores. However, the 10 CMIs to improve test scores are not exactly the same as the ones for increasing enrolment which prevents us from inferring policy conclusion regarding which outcomes are better incentives through CMIs.



Table 15: Summary of Effects of CMLs

Primary Outcome	Variable definition	Number of Interventions	Effect size	95% Confidence Interval	
Forensic economic estimates of corruption		3	0.15 (SMD)*	0.01	0.29
		2	0.08(RD)*	0.02	0.13
Perception measures of corruption		2	-0.23 (SMD)*	-0.38	-0.07
Access to service	Utilization (short term)	2	0.99 (ASE)	-1.05	3.02
	Utilization (medium term)	1	0.34 (ASE)*	0.12	0.55
	Immunization (short term)	2	1.56 (RR)*	1.39	1.73
	Immunization (medium term)	1	1.04 (RR)	0.52	2.61
	Enrolment	6	0.09 (SMD)	-0.03	0.21
	Dropout rate	7	-0.00 (SMD)	-0.10	0.10
Improvement in prevalence condition	Child death (short term)	2	0.76 (RR)	0.42	1.11
	Child death (medium term)	1	0.79 (RR)	0.57	1.08
	Weight for age (short term)	2	1.20 (RR)*	1.02	1.38
	Weight for age (medium term)	1	1.29 (RR)*	1.01	1.64
	Test score	10	0.16 (SMD)*	0.04	0.29
Quality of service	Average waiting time to get the service (short term)	2	0.99 (RR)	0.80	1.17
	Average waiting time to get the service (medium term)	1	1.06 (RR)	0.95	1.19

Note: \* Statistically significant at 95% confidence level.

The evidence suggests health conditions improved for child nutrition, as measured by weight for age in two studies, but not child mortality (CMLs are estimated to reduce mortality rates but the studies are underpowered to detect significant changes). The evidence also does not support reductions in waiting time as a result of CMLs.

Few papers assess the effect of CMLs on any of these outcomes, so our ability to generalize the summative findings of the review is limited.

The scant evidence found on mechanisms suggests that interventions that have modified intermediate outcomes, such as citizen participation in monitoring activities and providers and politicians' performance, have been those that include a set of tools for citizens to monitor providers or politicians, and facilitate contacts between both parts. These interventions appear to be the ones that have the bigger impacts on providers' responsiveness (lower absence rates, more teachers' effort, better school inputs) and more participation of communities in monitoring activities (more time spent in monitoring, more in-kind and monetary donations).

There are many reasons why interventions may fail to increase citizen participation in monitoring activities, such as limited information, or low expectations of leaders, officials, or service provider accountability or about the chances of success. In addition, there may be collective action failure, where some citizens free-ride on others' efforts to monitor the project. Certain groups, especially the poor, are less likely to participate in monitoring activities because they have more pressing priorities.

From the evidence included in this review we find that in the cases where citizens decided not to monitor service provision, providers did not change their behaviour. However, for interventions to increase politician and provider performance, hence improving service delivery, providers and politicians need to be accountable to citizens. Where enforcement mechanisms are lacking (e.g. if citizen support is not needed for politicians to stay in power) it is likely that CMLs will not improve outcomes along the causal chain. If, on the other hand, citizens can impose sanctions on unresponsive providers, CMLs are more likely to improve providers' performance.

Several studies emphasize the need to adapt interventions to local contexts including Björkman and Svensson (2010), Mansuri and Rao (2012) and Devarajan et al. (2011, 2013).

## **8. Conclusions**

This review identified and analyzed available evidence regarding the effects of CMIs on both access to and quality of service delivery and on corruption outcome measures in LMICs. The results suggest CMIs can have positive effects on corruption measures and some service delivery measures. However, due to the small number of available studies, results should be interpreted carefully. Interventions considered to address review question (1) may be not representative since they took place mainly in Africa and Asia, in rural communities within specific contexts, so the same interventions may have different effects elsewhere.

Further studies examining the effects of CMIs in different contexts are needed, especially factorial studies that are able to distinguish between different mechanisms of effect.

## References to included studies

Afridi, F. & Iversen, V. (2013). Social audits and MGNREGA delivery: Lessons from Andhra Pradesh, Brookings-NCAER India Policy Forum (eds. Barry Bosworth, Arvind Panagariya and Shekhar Shah).

Andrabi, T., Das, J. And Khwaja, A. I. (2013). Report Cards: The impact of Providing School and Child Test Scores on Educational Markets, Working paper JPAL.

Banerjee, A., Banerji, R., Duflo, E., Glennerster, R., Kenniston, D., Khemani, S., Shotland, M. (2007). Can Information Campaigns Raise Awareness and Local Participation in Primary Education? *Economic and Political Weekly*, 42(15): 1365-1372.

Banerjee, A. V., Banerji, R., Duflo, E., Glennerster, R., & Khemani, S. (2010). Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India. *American Economic Journal: Economic Policy*, 2(1), 1-30.

Barr, A., Mugisha, F., Serneels, P. and Zeitlin, A. (2012). Information and collective action in community-based monitoring of schools: Field and lab experimental evidence from Uganda. Working paper mimeo.

Björkman, M. & Svensson, J. (2009). Power to the people: evidence from a randomized field experiment on community-based monitoring in Uganda. *The Quarterly Journal of Economics*, 124(2), 735-769.

Björkman, M. & Svensson, J. (2010). When is community-based monitoring effective? Evidence from a randomized experiment in primary health in Uganda. *Journal of the European Economic Association*, vol. 8, issue 2-3, pages 571-581.

Björkman, M., de Walque, D. & Svensson, J. (2013) Information is Power: Experimental Evidence of the Long Run Impact of Community Based Monitoring, unpublished.

Gertler, P., Patrinos, H. A., & Rubio-Codina, M. (2008). Empowering parents to improve education: evidence from rural Mexico. Policy Research Working Paper 3935-IE, Revised May 2008.

Keefer, P., & Khemani, S. (2011). Mass media and public services: The effects of radio access on public education in Benin. *Policy Research Working Paper* Number 5559, Development Research Group, The World Bank.

Molina, E. (2013b). Bottom up institutional reform: Evaluating the impact of the citizen visible audit program in Colombia. Unpublished document.

Olken, B. A. (2004). Monitoring corruption: Evidence from a field experiment in Indonesia (No. w11753). National Bureau of Economic Research.

Olken, B. A. (2005). Corruption perceptions vs. corruption reality. (No. W12428). National Bureau of Economic Research

Olken, B. A. (2007). Monitoring corruption: evidence from a field experiment in Indonesia. *Journal of Political Economy*, 115(2).

Pandey, P., Goyal, S., & Sundararaman, V. (2009). Community participation in public schools: impact of information campaigns in three Indian states. *Education Economics*, 17(3), 355-375.

Pandey, P., Sehgal, A. R., Riboud, M., Levine, D., & Goyal, M. (2007). Informing Resource-Poor Populations and the Delivery of Entitled Health and Social Services in

Rural India. *JAMA: The Journal of the American Medical Association*, 298(16), 1867-1875.

Piper, B., & Korda, M. (2010). Early Grade Reading Assessment (EGRA) Plus: Liberia. Program evaluation report: Prepared for USAID/Liberia. Research Triangle Park, NC: RTI International.

Pradhan, M., Suryadarma, D., Beatty, A., Wong, M., Gaduh, A., Alisjahbana, A., & Artha, R. P. (2014). Improving educational quality through enhancing community participation: Results from a randomized field experiment in Indonesia. *American Economic Journal: Applied Economics*, 6(2), 105-126.

Reinikka, R., & Svensson, J. (2011). The power of information in public services: Evidence from education in Uganda. *Journal of Public Economics*, Elsevier, vol. 95(7), pages 956-966.

Singh, R., & Vutukuru, V. (2010). Enhancing Accountability in Public Service Delivery through Social Audits: A Case Study of Andhra Pradesh, India. Accountability Initiative, Centre for Policy Research, New Delhi.

Woodhouse, A. (2005) Village Corruption in Indonesia. Fighting Corruption in Indonesia's Kecamatan Development Program, World Bank Working Papers Series: Indonesian Social Development Paper No. 6.

### **Additional references cited**

Acemoglu, D. & J. A. Robinson. (2008). Persistence of Power, Elites, and Institutions. *American Economic Review* 98(1):267-293.

Banerjee, A. V. & Duflo, E. (2011). Poor Economics, a radical rethinking of the way to fight global poverty, Public Affairs.

Banerjee, A. V., & Mullainathan, S. (2008). Limited Attention and Income Distribution. *American Economic Review*, 98(2): 489-93.

Bardhan, P. (2002). Decentralization of governance and development. *Journal of Economic Perspectives*, 185-205.

Bardhan, P., & Mookherjee, D. (2006). Decentralisation and accountability in infrastructure delivery in developing countries. *The Economic Journal*, 116(508), 101-127.

Besley, T. and Persson, T. (2011). *Pillars of prosperity: The political economics of development clusters*. Princeton University Press.

Bhatnagar, D., Dewan, A., Moreno Torres, M., & Kanungo, P. (2003). Citizens' report cards on public services: Bangalore, India. *Empowerment Case Studies*. Available at: <http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTEMPOWERMENT/0,,contentMDK:20269087~pagePK:210058~piPK:210062~theSitePK:486411~isCURL:Y,00.html>

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley.

Booth, A. (2011). Searching for Studies. In: J. Noyes, A. Booth, K. Hannes, A. Harden, J. Harris, S. Lewin, & C. Lockwood (Eds). *Supplementary Guidance for Inclusion of Qualitative Research in Cochrane Systematic Reviews of Interventions*. Version 1 (updated August 2011). Cochrane Collaboration Qualitative Methods Group. Available at: <http://cqrmg.cochrane.org/supplemental-handbook-guidance>

Brinkerhoff, D.W., & Azfar, O. (2006). *Decentralization and community empowerment: Does community empowerment deepen democracy and improve service delivery?*. Washington, DC: Office of Democracy and Governance, USAID.

Campos, J. E., & Pradhan, S. (2007). *The Many Faces of Corruption : Tracking Vulnerabilities at the Sector Level*. ©World Bank, Washington DC. Available at: <https://openknowledge.worldbank.org/handle/10986/6848>

Capuno, J. J., & Garcia, M. M. (2010). Can information about local government performance induce civic participation? Evidence from the Philippines. *Journal of Development Studies*, 46(4), 624-643.

Casey, K., Glennerster, R., & Miguel, E. (2012). Reshaping institutions: Evidence on aid impacts using a pre-analysis plan. *Quarterly Journal of Economics*, 127(4):1755-1812.

Centre for Good Governance (2005). *Social audit: A toolkit. A guide for performance improvement and outcome measurement*. Centre for Good Governance, Hyderabad.

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., & Halsey Rogers, F. (2006). Missing in action: Teacher and health worker absence in developing countries. *Journal of Economic Perspectives*, Winter 2006, pp 91-116.

Chernozhukov, V. and Hansen, C. (2008). The Reduced Form: A Simple Approach to Inference with Weak Instruments. *Economics Letters*, 100(1), 68-71.

Devarajan, S., Khemani, S., & Walton, M. (2011). Civil society, public action and accountability in Africa. *HKS Faculty Research Working Paper Series* RWP11-036, John F. Kennedy School of Government, Harvard University



Egger, M., Davey Smith, G., Schneider, M. & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629-634.

Ferraz, C., Finan, F., & Moreira, D. B. (2012). Corrupting learning: Evidence from missing federal education funds in Brazil (Vol. w18150): National Bureau of Economic Research, Inc, NBER Working Papers: 18150.

Gaventa, J., & Barrett, G. (2012). Mapping the Outcomes of Citizen Engagement. *World Development* 40(12): 2399–2410.

Gorodnichenko, Y. & Sabirianova, P. K. (2007). Public Sector Pay and Corruption: Measuring Bribery from Micro Data. *Journal of Public Economics*, 91(5-6): 963-991.

Grandvoinnet, H., Aslam, G., & Raha, S. (2015). *Opening the Black Box: The Contextual Drivers of Social Accountability*. World Bank Publications.

Hanna, R., Bishop, S., Nadel, S., Scheffler, G., & Durlacher, K. (2011). *The effectiveness of anti-corruption policy: What has worked, what hasn't, and what we don't know*. DIFD Systematic Review.

Higgins, J., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions*. (Version 5.0.2, updated September 2009). The Cochrane Collaboration. Available at [www.cochrane-handbook.org](http://www.cochrane-handbook.org)

Keefer, P. & S. Khemani, S. (2004), "Why do the Poor Receive Poor Services?" *Economic and Political Weekly*, vol. 39, no. 9, pp. 935-943.

Keefer, P. & Khemani, S. (2005), "Democracy, Public Expenditures, and the Poor," *World Bank Research Observer*, vol. 20, no. 1, pp. 1-27.

Khemani, S. (2007). Can information campaigns overcome political obstacles to serving the poor? In S. Devarajan & I. Widlund (Eds.), *The politics of service delivery in democracies: Better access for the poor*. Expert Group on Development Issues, Ministry for Foreign Affairs, Sweden.

Khwaja, A. I. & Mian, A. (2005). Do Lenders Favor Politically Connected Firms? Rent Provision in an Emerging Financial Market, *The Quarterly Journal of Economics*, vol. 120(4), pages 1371-1411, November.

King, E., Samii, C. & Snilstveit, B. (2010): Interventions to promote social cohesion in sub-Saharan Africa, *Journal of Development Effectiveness*, 2(3), 336-370.

Kling, J. R., Liebman, J. B., Katz, L. F. & Sanbonmatsu, L. (2004) Moving To Opportunity and Tranquility: Neighborhood Effects on Adult Economic Self-Sufficiency and Health from a Randomized Housing Voucher Experiment, Princeton University Working Paper No. 5.

Lieberman, E., Posner, D. N. & Tsai, L. (2013). Does Information Lead to More Active Citizenship? Evidence from an Education Intervention in Rural Kenya. Technical report MIT Political Science Department.

Mansuri, G., & Rao, V. (2012). *Localizing development: Does participation work?* Washington DC, World Bank.

Maru, V. (2010). Social accountability and legal empowerment. *Health and Human Rights: An International Journal*, North America, 1225 05.

Mauro, P. (1995). Corruption and Growth. *The Quarterly Journal of Economics*, 110(3): 681-712.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods*, 1(1), 30.

McMillan, J. & Zoido, P. (2004). How to Subvert Democracy: Montesinos in Peru. *Journal of Economic Perspectives*, 18(4): 69-92.

Molina, E. (2013a). Community monitoring and self-fulfilling prophecies in service delivery. Unpublished Manuscript.

Molina, E., Carella, L., Pacheco, A., Cruces, G., & Gasparini L. (2016). Community monitoring interventions to curb corruption and increase access and quality of service delivery in low- and middle-income countries. *Campbell Systematic Reviews* 2016:8 DOI: 10.4073/ csr.2016.8

Nunn, N., & Wantchekon, L. (2009). The slave trade and the origins of mistrust in Africa (No. w14783). National Bureau of Economic Research.

Olken, B. A. (2006). Corruption and the Costs of Redistribution: Micro Evidence from Indonesia. *Journal of Public Economics*, 90(4-5):853-870.

Olken, B. A. (2009). Corruption Perceptions vs. Corruption Reality. *Journal of Public Economics*, 93(7-8): 950-964.

Olken, B. A. & Barron, P. (2009). The Simple Economics of Extortion: Evidence from Trucking in Aceh. *Journal of Political Economy*, 117(3): 417-452.

Olken, B. A., & Pande, R. (2012). Corruption in developing countries. *Annual Review of Economics*, 4, 479-509.

Olson, M. (1971). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard economic studies, v. 124 revised ed. Harvard University Press.

Pan, L., & Christiaensen, L. (2012). Who is vouching for the input voucher? Decentralized targeting and elite capture in Tanzania. *World Development*, Elsevier, vol. 40(8), pages 1619-1633.

Pande, R. & Olken, B. A. (2011). *Governance review paper*. JPAL governance initiative. Abdul Latif Jameel Poverty Action Lab.

Pandey, P. (2010). Service delivery and corruption in public services: How does history matter? *American Economic Journal: Applied Economics*, 2(3), 190-204.

Persson, T., & Tabellini, G. (2002). *Political Economics: Explaining Economic Policy*. Vol. 1 of MIT Press Books The MIT Press.

Reinikka, R., & Svensson, J. (2004). Local capture: evidence from a central government transfer program in Uganda. *The Quarterly Journal of Economics*, 119(2), 679-705.

Reinikka, R., & Svensson, J. (2005). Fighting corruption to improve schooling: Evidence from a newspaper campaign in Uganda. *Journal of the European Economic Association*, MIT Press, vol. 3(2-3), pages 259-267, 04/05.

Reinikka, R., & Svensson, J. (2011). The power of information in public services: Evidence from education in Uganda. *Journal of Public Economics*, Elsevier, vol. 95(7), pages 956-966.

Remme, J. H. F. (2010). Community-directed interventions for priority health problems in Africa: results of a multicountry study. *Bulletin of the World Health Organization*, 88(7), 509-518.

Ringold, D., Holla, A., Koziol, M., & Srinivasan, S. (2012). *Citizens and service delivery. Assessing the use of social accountability approaches in the human development sectors*. Washington, DC: World Bank.

Rose-Ackerman, S. (2004). *Governance and Corruption*, in Global Crises, Global Solutions. B. Lomborg, ed. Cambridge: Cambridge University.

Sacks, A. & Larizza, M. (2012). Why Quality Matters: Rebuilding Trustworthy Local Government in Post-Conflict Sierra Leone. World Bank Policy Research Working Paper No.

Shadish, W. & Myers, D. (2004). Research design policy brief. Campbell

Collaboration: Oslo. Available at:

[http://www.campbellcollaboration.org/artman2/uploads/1/C2\\_Research\\_Design\\_Policy\\_Brief-2.pdf](http://www.campbellcollaboration.org/artman2/uploads/1/C2_Research_Design_Policy_Brief-2.pdf)

Singer, M. M. (2013). Bribery Diminishes Life Satisfaction in the Americas, Working paper from University of Connecticut. Unpublished manuscript.

Snilstveit, B. (2012). Systematic reviews: from 'bare bones' reviews to policy relevance. *Journal of Development Effectiveness*, 4(3), 388-408.

Stiglitz, J. E. (2002). Participation and development: Perspectives from the comprehensive development paradigm. *Review of Development Economics*, 6(2), 163-182.

Sukhtankar, S. (2011). Sweetening the Deal? Political Connections and Sugar Mills in India American Economic Journal: Applied Economics, vol 4, no 3, pp 43-63, July 2012.

Svensson, J. (2003). Who Must Pay Bribes and How Much? Evidence from A Cross Section Of Firms. *The Quarterly Journal of Economics*, 118(1): 207-230.

Svensson, J. (2005). Eight questions about Corruption. *Journal of Economic Perspectives*, 19 (5): 19-42.

Tosi, F. G. (2012). Direct popular participation and crises in public services in Argentina: The Gordian Knot. Rochester.

Transparency International. (2013). *Policy Brief. Looking Beyond 2015: A role for governance*. Transparency International, Berlin, Germany.

UN (2008). *People matter: Civic engagement in public governance*. New York, NY: United Nations Department of Economic and Social Affairs.

Waddington, H., White, H., Snilstveit, B., Hombrados, J. G., Vojtkova, M., Davies, P., & Tugwell, P. (2012). How to do a good systematic review of effects in international development: a tool kit. *Journal of development effectiveness*, 4(3), 359-387.

Waddington, H, Snilstveit, B, Hombrados, J, Vojtkova, M, Phillips, D, Davies, P & White, H. (2014) Farmer Field Schools for Improving Farming Practices and Farmer Outcomes: A Systematic Review *Campbell Systematic Reviews* 2014:6 DOI: 10.4073/csr.2014.

White, H. (2009). Theory-based impact evaluation: Principles and practice. Working Paper. International Initiative for Impact Evaluation: New Delhi. Available at [http://www.3ieimpact.org/admin/pdfs\\_papers/51.pdf](http://www.3ieimpact.org/admin/pdfs_papers/51.pdf)

World Bank. (2003). *Making services work for poor people. World Development Report 2004*. Washington, DC: World Bank; New York: Oxford University Press.

Zitzewitz, E. (2012). Forensic Economics. *Journal of Economic Literature*, 50(3): 731-69.

## Appendix

Table A1: Description of included studies

Study (sibling study)	Location and treatment unit	Intervention	Target and comparison group	Study design / attribution method	Outcomes data collected	Follow-up time period <sup>1</sup>	Internal validity assessment
Afridi, F. and Iversen, V. (2013)  (Singh and Vutukuru, 2009)	India: Andhra Pradesh. Gram Panchayats	Social audit (Second audit)  Social audit (Third audit)	This is a panel data set which comprises official data from three rounds of social audits, with an initial sample of 300 GPs in 8 districts. It compares the results of the second (264 audits) and third audit (166) with those of the first one (284). 548 number of audits, from which 284 are first audit, and 264 are a second audit. (Clusters at GPs level: 300). The intervention is the Mahatma Gandhi National Rural Employment Guarantee Schemes (MGNREGA).	Difference-in- differences (DID)	Perception of corruption: number of irregularities	5 years	Low risk of bias
Andrabi, Das and Khwaja (2013)	Pakistan. Schools and children	Scorecard	Villages were sampled from 3 districts: one each in the north, center and south. Treatment group: Scorecards. Control Group without scorecard 112 Villages were chosen randomly from among those with at least one private school according to a 2000 census of private schools. First, Grade 3 children in all primary schools were tested and then, in a randomly selected 50 percent of the villages, were disseminated report cards with the results of school and child test scores for all schools (804) and tested children (12110)	Difference-in- differences (DID)	Access to service: education enrollment Changes in prevalence: test scores	1 year and 2 years	Low risk of bias
Banerjee <i>et al.</i> (2010)	India: Uttar Pradesh. Villages	Information campaign (IC) Treatment 1	The evaluation took place in 280 villages in Jaunpur district. 85 villages as control group and 195 as target group. The final sample for the baseline survey	RCT	Access to service: education enrollment Changes in prevalence: test scores	1 year	Low risk of bias





Study (sibling study)	Location and treatment unit	Intervention	Target and comparison group	Study design / attribution method	Outcomes data collected	Follow-up time period <sup>1</sup>	Internal validity assessment
	was the same as described in Björkman and Svensson (2009).			regression (Kling et al., 2004)			
Gertler <i>et al.</i> (2008)	Mexico. Non- indigenous primary schools in rural areas	Scorecard	Treatment schools are those schools that first received the AGE program at the beginning of any school year between 1998-99 and 2001- 02, and had AGE continuously ever since (N=2544). Those that had not received AGE before school year 2002-03 constitute the comparison group (N=3483).	Difference-in- differences (DID)	Access to services: education enrollment and repetition	12 years	Low risk of bias
Keefer and Khemani (2011)	Benin. Households and children in second grade	Information Campaign Intervention (IC)	In the target group are the households and children in the villages which access to the radio. In the control are those in villages without access to the radio. 210 villages (4200 households) from 21 communes	Cross-section, Quasi- experimental	Changes in prevalence: test scores	Not applicable	Low risk of bias
Molina (2013b)	Colombia. Projects and households	Social audit	The random sample contains 390 households for the 13 projects in the treatment group and 410 for the 11 projects in the control group.	Cross section- Matching	Perception of corruption: perception of adequacy in administration of resources	Not applicable	Low risk of bias
Olken (2007) (Olken, 2004, 2005)	Indonesia. Villages	Social Audit: Invitations	Social Audit with Invitations vs. Social Audit. 199 villages (audit 94 , control 105)	RCT, Cross- section (regression)	Forensic estimates of corruption: percent missing funds major items in roads and ancillary projects	1 year	Low risk of bias
(Woodhouse, 2005)		Social Audit: Invitations + comments	Social audit with invitations plus comments vs. Social Audit 202 villages (audit 96, control 106)				

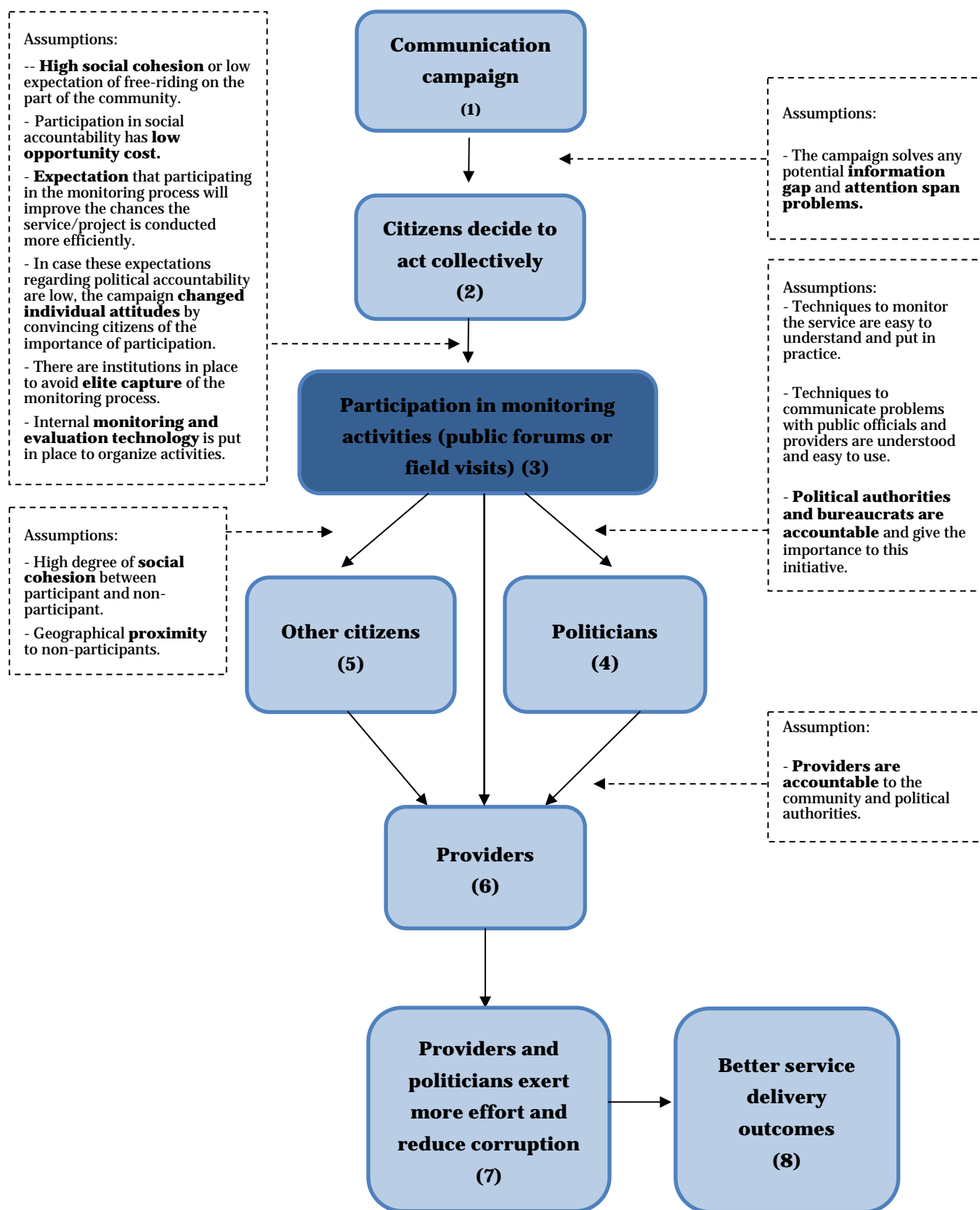
Study (sibling study)	Location and treatment unit	Intervention	Target and comparison group	Study design / attribution method	Outcomes data collected	Follow-up time period <sup>1</sup>	Internal validity assessment
Pandey <i>et al.</i> (2007)	India: Uttar Pradesh. Households	Information campaign (IC)	105 villages (1045 households at the baseline), from which 55 (548 households) intervention and 50 (497 households) control	Difference-in-differences (DID) and Cross-section	Perception of corruption: development work in villages Access to service: Visits by nurse midwife; prenatal examinations, tetanus vaccinations, and prenatal supplements received by pregnant women; vaccinations received by infants.	1 year	Low risk of bias
Pandey, Goyal and Sundararaman (2009)	India: Uttar Pradesh, Madhya Pradesh and Karnataka. Teachers and students in villages	Information campaign (IC)  Second IC in one region	610 villages from which 340 intervention and 270 control	RCT	Changes in prevalence: test scores	2-4 months	Medium risk
Piper and Korda (2010)	Liberia. Schools	Information campaign (IC)	Groups were randomly selected and clustered within districts, such that several nearby schools were organized together. 117 schools, from which 59 are control. The intervention was targeted at grades 2 and 3.	RCT	Changes in prevalence: test scores	1 year	Medium risk
Pradhan <i>et al.</i> (2014)	Indonesia. Schools	Training (T): IC  Linkage (L): IC	2 provinces, 9 districts, 44 subdistricts and 520 schools. Training: treatment group: 230 schools, 1060 students; control group 190 schools and 2120 students. Linkage: treatment group: 240 schools and 893 students; control group: 180 schools and 2120 students. The authors also include a third	Difference-in-differences (DID)	Changes in prevalence: test scores	1 year and 10 months	Low risk of bias

Study (sibling study)	Location and treatment unit	Intervention	Target and comparison group	Study design / attribution method	Outcomes data collected	Follow-up time period <sup>1</sup>	Internal validity assessment
		The authors also include a third treatment that we do not consider as it is not an eligible CMI for this review	treatment that we do not consider as it is not of the type of CMI considered in this review, the intervention introduced changes in the election of the committee. They also explore combinations of treatments given that some individuals in the control groups for each treatment had received the other treatments.				
Reinikka and Svensson (2011)	Uganda. Schools	Information campaign (IC)	Using distance to newspapers outlets the authors construct the treatment and control group. 218 schools for which survey data are available for the years 1991-95 and 2001, and a sample of 388 (218 + 170) schools for which survey data are available in 2001.	Difference-in-differences (DID), Instrumental Variable (IV)	Forensic economic estimates of corruption: share of grants received Access to service: education enrollment Changes in prevalence: test scores	Not applicable	Low risk of bias

Notes: IC Information campaign, SA Social Audit; SC Scorecard; DID: Differences-in-differences, IV: Instrumental Variable; OLS: Ordinary Least Squares estimation.

1/ Average years from intervention to endline survey.

Figure A1: Revised causal chain with assumptions identified from evidence



---

<sup>i</sup> There is also a new Global Partnership for Social Accountability (GPSA), which is a coalition of donors, governments and civil society organisations aiming to improve development results by supporting capacity building for enhanced citizen feedback and participation to monitor service delivery. GPSA aims to reach overall funding of 75 to 125 million USD over the next seven years. To date, 15 countries have joined the GPSA: Bangladesh, Belarus, Colombia, Dominican Republic, Honduras, Indonesia, Kyrgyzstan, Malawi, Moldova, Mongolia, Mozambique, Philippines, Senegal, Tajikistan and Tunisia.

<sup>ii</sup> For the purpose of this review, service delivery involves not only services, but also construction of necessary infrastructure to carry out those services. As a result, we will talk indistinctly between service delivery and project performance.

<sup>iii</sup> The term bottlenecks has been used in the literature (Lieberman et al., 2013) to refer to constraints that limit the effectiveness of community monitoring programmes.

<sup>iv</sup> In order to give salience to information practitioners use an array of instruments to attract the citizens' attention. We are not aware of any CMI where these incentives were embedded in the theory of change and properly assessed. This appears to be a knowledge gap for CMIs.

<sup>v</sup> In some parts of the world the state fails completely to provide services and to monitor illegal private service provision. Even under these environments, when citizens can choose providers overall providers' performance may increase.

<sup>vi</sup> We also aimed to answer a third review question – “to what extent do geographic region, income level or length of exposure to interventions explain the variation in findings?” – using moderator analysis. Due to the limited number of studies, we were unable to perform planned moderator analyses by study design and length of exposure to intervention. Moderator analysis by region did not reveal big differences.

<sup>vii</sup> We critically appraised the consistency among ratings by our coders through inter-rater assessment (McGraw and Wong, 1996). The absolute agreement intra-class correlation is 0.70 with a 95% CI [0.21, 0.95]. In case of disagreement, a third researcher determined the rating.

<sup>viii</sup> ASE (Kling *et al.*, 2004) is interpreted in the same way as SMD. It is calculated by combining several measures for the same outcome into a unique average standardized treatment effect (ASE), by estimating a seemingly unrelated regression system for  $K$  related outcomes:  $Y = [I_K \otimes (T \ X)] \theta + v$ , where  $I_K$  is a  $K$  by  $K$  identity matrix. The average standardized treatment effect is estimated as

$$\tilde{\beta} = \frac{1}{K} \sum_{k=1}^K \frac{\tilde{\beta}_k}{\tilde{\sigma}_k},$$
 where  $\tilde{\beta}_k$  is the point estimate on the treatment indicator in the  $k^{th}$  outcome regression and  $\tilde{\sigma}_k$  is the standard deviation of the control group for outcome  $k$  (Björkman de Walque and Svensson, 2013).

<sup>ix</sup> In some cases, these correlations were available in the studies' databases, or where easily obtainable from tables reported in the papers. When not available, we assumed a value of 0.5, and checked whether the results changed substantially for extreme correlation values.

<sup>x</sup> All randomized field experiments report no statistical difference between treatment and control groups.

<sup>xi</sup> In both cases, we changed the sign of the effect size so it can be interpreted properly (that is, a positive effect size means that corruption has been reduced).

<sup>xii</sup> The study also evaluates the effect of external audits, which did reduced corruption, but we did not include it in the meta-analysis because it does not fall into any of our four intervention categories.

<sup>xiii</sup> This finding is consistent with those reported by the author, who argues that ‘increasing grassroots participation in monitoring had little average impact (Olken, 2007).

<sup>xiv</sup> Finally, we did not include Banerjee *et al.* (2010) as it is not a measure of corruption they use, but rather they look at whether the treatments to increase community monitoring generated additional nonteaching resources for the schools. They found that none of the interventions have any effect.

<sup>xv</sup> Actually, the medium term impact of the first intervention is assessed in Björkman, de Walque and Svensson (2013). However, to avoid confusion, we designate the latter as the main reference for the second intervention and Björkman and Svensson (2009) for the first intervention.

<sup>xvi</sup> We were not able to compute neither SMD nor RR for these outcomes due to lack of information.

<sup>xvii</sup> The short term impact of this intervention is also not statistically significant, but it is not reported in the table since we were not able to compute RR.

<sup>xviii</sup> It is important to note why we think this is a quality measure and not an access measure. Access is related to getting the service. However, you can get the service and the fact that you had to wait makes it less valuable and of lesser quality.

<sup>xix</sup> When different test scores were reported (e.g. language and math test scores), we previously pooled them following the procedure explained before.

<sup>xx</sup> It should be noted that we are excluding two studies for which we were not able to compute standardized effects (Table 12) but which found significant effects of CMIs on our outcomes of

---

interest. Keefer and Khemani (2011) examine the proportion of children tested in the village public school who could read sentence and paragraphs (ASER literacy test), finding that the information campaign resulting from communities' access to radios enhanced literacy tests. Pandey, Goyal and Sundararaman (2009) examined the percentage of children who could pass different learning tests, including reading and writing competences and mathematics abilities.

<sup>xxi</sup> Additional qualitative and quantitative evidence support this. For example, Gaventa and Barrett (2012) perform a meta-case study of 100 interventions aim to increase citizen engagement in service delivery. For the 828 outcomes from the 100 reviewed case studies, only 153 came from interventions where the final goal was to strengthen the responsiveness and accountability of the state to provide services. Results indicate that 55% of those 153 outcomes were positive and 45% were negative. Negative results were associated with failure of citizens to participate, due in part to fear of backlash against those who speak out and a sense of tokenism in the participation mechanism.