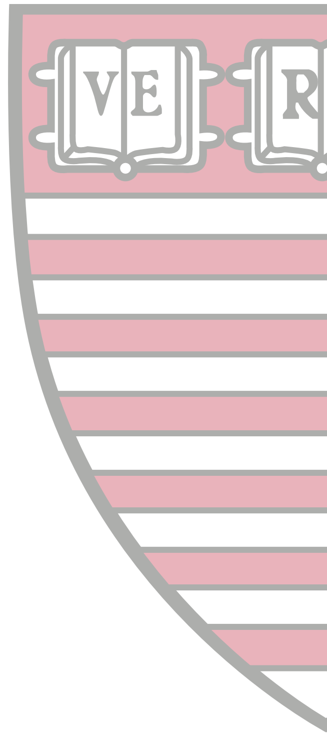
OCCASIONAL PAPERS SERIES



Brokering Collaboration: Involving Officials in Community Scorecard Programs

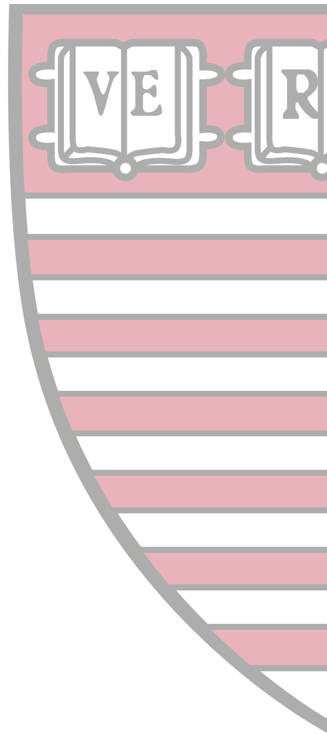
Stephen Kosack, Jessica Creighton, Courtney Tolmie, Fatu Conteh, Eric Englin, Linda Gassama, Hannah Hilligoss, Takondwa Iman, Mitchris Chapman Kodam, Agnes Lahai, James Rasaiah, Emmanuel Sapalado, J. Preston Whitt, James Quophy Wumenu, Archon Fung

Tony Saich, *Series Editor*

April 2021

HARVARD Kennedy School
ASH CENTER
for Democratic Governance
and Innovation

OCCASIONAL PAPERS SERIES



Brokering Collaboration: Involving Officials in Community Scorecard Programs

Stephen Kosack, Jessica Creighton, Courtney Tolmie, Fatu Conteh, Eric Englin, Linda Gassama, Hannah Hilligoss, Takondwa Iman, Mitchris Chapman Kodam, Agnes Lahai, James Rasaiah, Emmanuel Sapalado, J. Preston Whitt, James Quophy Wumenu, Archon Fung

Tony Saich, *Series Editor* April 2021

About the Ash Center

The Roy and Lila Ash Center for Democratic Governance and Innovation advances excellence and innovation in governance and public policy through research, education, and public discussion. By training the very best leaders, developing powerful new ideas, and disseminating innovative solutions and institutional reforms, the Center's goal is to meet the profound challenges facing the world's citizens. The Ford Foundation is a founding donor of the Center. Additional information about the Ash Center is available at ash.harvard.edu.

This research paper is one in a series published by the Ash Center for Democratic Governance and Innovation at Harvard University's John F. Kennedy School of Government. The views expressed in the Ash Center Policy Occasional Paper Series are those of the author(s) and do not necessarily reflect those of the John F. Kennedy School of Government or of Harvard University. The papers in this series are intended to elicit feedback and to encourage debate on important public policy challenges

This paper is copyrighted by the author(s). It cannot be reproduced or reused without permission. Pursuant to the Ash Center's Open Access Policy, this paper is available to the public at ash.harvard.edu free of charge.

A PUBLICATION OF THE

Ash Center for Democratic Governance and Innovation
Harvard Kennedy School
79 John F. Kennedy Street
Cambridge, MA 02138

617-495-0557
ash.harvard.edu

Abstract

Programs to improve the transparency and accountability of public services are an increasing focus of international commitments to sustainable development. We ask whether involving officials in one common approach—community scorecard programs—brokers state-society collaboration that improves public services. We compare two scorecard programs focused on improving maternal and newborn health care that were offered in 215 communities similarly stratified across five countries. The first program, offered in 200 communities in Indonesia and Tanzania, involved facilitated meetings among community members. A similar program in 15 communities in Ghana, Malawi, and Sierra Leone involved facilitated meetings among community participants as well as between community members and hereditary authorities (in Malawi) or district-level elected and appointed officials (in Ghana and Sierra Leone). Interviews, focus groups, and systematic observations consistently suggest that in the program in Malawi, participants took similar approaches to improving their health care to participants in Indonesia and Tanzania—focusing primarily on improving care themselves and with health-care providers and others in their communities—and that the results of their efforts were similar to the program in Indonesia and Tanzania, where a randomized controlled impact evaluation found that average community outcomes did not improve significantly faster than in a control group of communities. In both Ghana and Sierra Leone, participants collaborated more with officials and saw tangible changes to health care that they and others noticed and remembered in nearly twice the proportion of communities as in the program in Indonesia and Tanzania. We conclude that involving officials in these programs may increase their effectiveness.

Acknowledgments

We are saddened by the passing of our coauthor Linda Gassama and our colleagues Jessica Engelman, Chris Grorud, and Helen Richards, and gratefully acknowledge their many contributions to this work. We thank Daniel Berliner, Akshay Dixit, Michael Jarvis, Patricia Scheid, and Musa Ansumana Soko for helpful comments on earlier drafts and Elizabeth Foz for careful and helpful comments and editing. This paper stems from the Transparency for Development project, a multicountry mixed-method research project that seeks generalizable knowledge about the impacts and implications of transparency and accountability programs. It builds on eight years of collaboration with colleagues at PATTIRO in Indonesia, the Clinton Health Access Initiative in Tanzania, the Center for Democratic Development in Ghana, the Malawi Economic Justice Network (MEJN), the Water, Sanitation and Hygiene Network (WASH-Net) in Sierra Leone, the Results for Development Institute, the Evans School of Public Policy and Governance at the University of Washington, and the Ash Center for Democratic Governance and Innovation at the Harvard Kennedy School. Over the years we have drawn insight, ideas, and inspiration from many scholars and practitioners in the fields of transparency and accountability and social accountability and related disciplines, among them Yamini Aiyar, Rikhil Bhavnani, Kevin Bohrer, Nick Burnett, Jessica Cohen, Jishnu Das, Jonathan Fox, Rachel Glennerster, Mary Graham, Florencia Guerzovich, Scott Guggenheim, Hahrie Han, Nathaniel Heller, Patrick Heller, Vanessa Herringshaw, Michael Jarvis, Anuradha Joshi, Asim Khwaja, Jenny Lah, Scott Mainwaring, Jane Mansbridge, Tom McEnroe, Jim McGuire, Nushina Mir, Ben Olken, Rohini Pande, Alfonsina Peñaloza, Enrique Peruzzotti, Dan Posner, Lant Pritchett, the late Helen Richards, Susan Rose-Ackerman, Bill Savedoff, Patricia Scheid, Jakob Svensson, Lily Tsai, Michael Walton, David Weil, Tom Wingfield, and Michael Woolcock; our co-principal investigators, Jean Arkedis and Dan Levy; and our current and former colleagues on the Transparency for Development team—Sarah Alphs, Novi Anggraeni, Iqra Anugrah, Rexford Asiama, Mohammed Awal, Grant Bridgman, Kelvin Chirwa, Megan Cogburn, Sad Dian Utomo, Akshay Dixit, the late Chris Grorud, Riska Hasan, Jenna Juwono, June Kambalametore, Kondwani Kaunda, Hilda Kigola, Esther Mtumbuka, Rohit Naimpally, Happy Ndomba, Didik Purwandanu, Mohamed Yunus Rafiq, Lindsey Roots, Rachmat Reksa Samudra, Emmanuel Sandy, Manpreet Singh, Musa Ansumana Soko, Astri Arini Waluyo, and Kankan Xie. Our thanks as well to Arja Dayal and colleagues at Innovations for Poverty Action in Sierra Leone, Lina Marliani and colleagues at J-PAL Southeast Asia, Wayan Suriastini and her team at SurveyMETER in Indonesia, and Samson Kiware, Isaac Lyatuu, Diego Shirima, and the rest of their team at Ideas in Action in Tanzania for superb data collection. We are grateful to seminar participants at Brown, Harvard, the Institute for Development Studies at the University of Sussex, the London School of Economics, and Oxford for helpful questions about and comments on earlier versions of this work. The study was made possible by generous financial support from the Bill & Melinda Gates Foundation, the William and Flora Hewlett Foundation, and the UK Department for International Development, under the coordination of the Transparency and Accountability Initiative. Most importantly, we gratefully acknowledge the thousands of Indonesians, Ghanaians, Malawians, Sierra Leoneans, and Tanzanians who participated in the two programs. Interview and observation protocols, replication scripts for the figures in this paper, and background documents and reports and other working papers on the broader project are available at t4d.ash.harvard.edu.

Contents

1. Introduction 1

2. Community Scorecard Programs and State Actors 2

3. Two Community Scorecard Programs in Five Countries 5

A. Improving Health Systems 7

B. Competitive Democratic Institutions 9

C. Two Community Scorecard Programs 10

4. Methods 13

5. Findings 17

A. Ghana and Sierra Leone: A Community Scorecard Program Involving District Officials 19

B. Malawi: A Community Scorecard Program Involving Traditional Authorities 27

6. Generalizability and Scope Conditions 29

7. Conclusion 31

References 33

Introduction to the Appendixes 39

Appendix A. Contexts 40

Appendix B. Two Community Scorecard Programs 44

Appendix C. Planned Activities 49

Appendix D. Perceptions of Efficacy 59

Appendix E. Responses by Officials 65

Appendix F. Anchoring Vignettes 67

Appendix References 69

Notes 71

Figures

Figure 1. Declining infant mortality rates, 1960–2018 8
Figure 2. Hypothesis in context 10
Figure 3. Six perspectives on participants' experiences 16
Figure 4. Engaging with officials 23
Figure 5. Efforts and outcomes 25
Figure 6. Perceptions of responsiveness 26

Appendix Figures

Figure A.1. Infant mortality rates in the year of the program (2015 or 2017) 40
Figure A.2. Infant mortality and democracy 42
Figure B.1. A scorecard from the program in Indonesia 45
Figure B.2. A scorecard meeting in Tanzania 45

Tables

Table 1. Co-creation: Four pathways to improved public services 4
Table 2. Five country contexts 6
Table 3. Observable implications of hypothesis across contexts 14
Table 4. Planning, perceptions of efficacy, and later reflections on participating 18
Table 5. Planned activities to improve health care 19
Table 6. Objectives of participants' plans 20
Table 7. Results expected and observed 29

Appendix Tables

Table C.1. Number of activities planned 49
Table C.2. Examples of participants' plans 50
Table C.3. Primary actors involved in all planned activities 54
Table C.4. Primary actors participants planned to involve in at least one activity 55
Table C.5. All actors participants planned to involve in at least one activity 56
Table C.6. Level of all actors participants planned to involve 57
Table C.7. Progress in activities 58
Table C.8. Completed activities across communities 58
Table D.1. Individual perceptions of civic efficacy 59
Table D.2. Skepticism among participants (meeting observations) 61
Table D.3. Reflections on participating 62
Table D.4. Changes participants recalled as a result of their efforts 63
Table E.1. Verified responses 65
Table F.1. Perceptions of responsiveness 67

1. Introduction

International development efforts increasingly include programmatic approaches to increasing the transparency and accountability of public services to the communities they serve (Fox 2007a, 2015; Joshi 2010; J-PAL 2011; McGee and Gaventa 2011; World Bank 2004).[1] The growth of these programs is one result of a decades-long commitment to sustainable development programs that are participatory and focused on expanding human capabilities, particularly through increasing access to quality education, health care, and other public services.[2] Transparency and accountability programs vary widely but typically provide information designed to make the performance of public services more transparent—such as test scores or health metrics, budgetary allocations, or information about whether clinics or schools are well equipped and staffed—and often encourage the communities who use those services to participate in varied ways in improving their responsiveness and effectiveness, such as choosing among available providers, collaborative problem-solving, or making complaints. A growing literature has come to mixed conclusions about the effectiveness of these programs. Several studies show measurable, even transformative improvements; others, minimal or no effects.[3]

In this paper, we ask whether involving officials and leaders in positions of influence and authority in one common approach, community scorecard programs, brokers state-society collaboration that contributes to more improvement in public services than when these programs do not involve these brokered connections with officials. Community scorecard programs sometimes involve regional or national officials, but typically they are local, community-focused programs that offer information about the performance of public services and facilitated forums designed to encourage community deliberation, planning, and problem-solving to improve those services.[4] We compare the experiences of participants in two community scorecard programs focused on improving the same public service—maternal and newborn health care—in randomly selected, mostly rural communities in five countries: Indonesia, Ghana, Malawi, Sierra Leone, and Tanzania. These five countries vary widely on many political, economic, social, cultural, and geographic dimensions but also share two relative similarities of potential relevance to how these programs may improve public services. First, maternal and newborn health-care systems in all five, while still underresourced and inaccessible to many, have been steadily improving for decades. Second, other than Tanzania, which was a competitive authoritarian regime at the time of the program, all shared political institutions that, although not fully democratic, had seen regular alternation of national political power between elected representatives of rival constituencies.

The first of the two programs was offered in 2015 and 2016 in 200 randomly selected communities similarly stratified across two provinces of Indonesia and two regions of Tanzania (100 communities in each country). In almost all communities where the program was offered, participation was substantial and sustained, and although some participants were discouraged and skeptical, most recalled their experiences as beneficial and helpful for improving their community's health care. In about one-third of communities, participants and others in their communities saw their efforts lead to noticeable and memorable changes. But in the average community, their efforts were not sufficient to improve community health or health-care outcomes further or faster than the average in a comparable group of communities stratified across the same regions and provinces where the program was not offered ($p > .05$) (Arkedis et al., forthcoming).

The second program, offered in 15 randomly selected communities in Ghana, Malawi, and Sierra Leone in 2017 and 2018 (in each country, five communities across one district), was similar in most respects, but facilitators also organized meetings between community members and district officials and leaders who had expressed interest in the program and willingness to participate in it: health, planning, and administrative officers, elected leaders, hereditary traditional leaders, and other district-level officials

and leaders in positions of influence and authority. This paper explores the implications across these varied country contexts of this difference in these otherwise similar community scorecard programs.

In brief, the evidence suggests that the second program involving leaders and officials with responsibility for and authority over district public services offered opportunities for mutually constructive collaboration between citizens and officials that contributed to more improvements in access to quality health care. As with the first program, in almost all the communities where the second program was offered a group of community members participated meaningfully—discussing information the facilitator provided on maternal and newborn health care, deliberating on what would improve it, and planning and attempting to improve it or convince others to—and generally recalled the experience to be helpful for improving their community's maternal and newborn health care. In Malawi, where facilitators also organized meetings with traditional authorities, participants' approaches and experiences were similar to those of participants in Indonesia and Tanzania in focusing mostly on what participants could do themselves, with others in their communities, and with health-care providers. But in Ghana and Sierra Leone, where the program involved district-level officials with more direct authority and responsibility over district public services, they and community members used the meetings to discuss how to improve understanding and resource allocations, clear up misperceptions and mistrust, and otherwise improve the communities' access to quality care. Some of the discussions and plans involved social accountability for officials' existing responsibilities; others involved community members, officials, or both taking on new tasks or responsibilities. Overall, these discussions and plans led to far more collaboration between community participants and officials than in Malawi, Indonesia, or Tanzania. Both community participants and officials involved generally recalled these collaborations as mutually constructive, and interviews with them as well as community observations suggest that although they did not lead to tangible changes in every community, they complemented participants' community-based efforts, leading to nearly twice as many communities seeing tangible changes in their health care as in Indonesia or Tanzania. These included changes that may have been challenging for community participants to realize without these collaborations with officials, such as a new water supply at one health facility, the opening of a new facility, and initial construction of another facility.

The paper proceeds as follows. Section 2 discusses the hypothesis that a community scorecard program involving officials can broker more collaboration between community members and officials that contributes to improvements in public services. Section 3 describes the two programs as well as the two similarities noted above in the otherwise varied country contexts where they were offered—in health-care systems and in political institutions—that are of potential relevance to the generalizability of our findings. Section 4 describes the methods we use to explore the hypothesis in Section 2. Section 5 details our findings. Section 6 discusses the generalizability and related scope conditions of the findings to other contexts and programs. Section 7 concludes.

2. Community Scorecard Programs and State Actors

Transparency and accountability programs vary widely in their involvement of state actors. In its 2004 *World Development Report*, the World Bank conceptualized a basic distinction between the "long route" of accountability, in which citizens work with officials who oversee clinics and schools, control resources, or exercise supervisory, regulatory, policymaking, or other authority to improve public services, and the "short route" of seeking improvements directly with service providers. Community scorecard programs tend to be relatively focused on the short route. These programs sometimes involve district and even national officials.[5] But among the broader range of programmatic approaches to improving transparency and accountability of public services, community scorecard programs are relatively local and focus

on problem-solving among community members, service providers, and local government rather than regional or national officials (Björkman and Svensson 2009; Björkman Nyqvist, de Walque, and Svensson 2017; Christensen et al. 2020; Gullo, Galavotti, and Altman 2016; Raffler, Posner, and Parkerson 2019).

Can involving officials and leaders in positions of influence and authority in these programs broker collaboration between community members and officials that contributes to more improvement in public services? Scholars have long argued that collaboration and co-production between officials and citizens can be mutually empowering and can help solve problems of principal-agent accountability in the systems that deliver public services, and this argument has been influential in programmatic approaches to transparency and accountability and social accountability.[6] Several well-known programs are creations of governments, such as participatory budgeting in Brazil (Baiocchi, Heller, and Silva 2011; Souza 2001; Wampler 2007), or are grassroots initiatives later adopted by governments, such as social audits in India (Aiyar, Mehta, and Samji 2011; Jenkins and Goetz 1999). Some scholars have also argued that effective state-society collaboration can involve more informal and short-term interactions encouraged and supported by external actors such as nongovernmental organizations and donors.[7] Several donor-supported programs involve close collaboration between government and communities, such as the Health Systems Strengthening Project in India (World Bank 2018), community-based monitoring and planning of health services in India (Marathe et al. 2020), and government audits and community monitoring of infrastructure projects in Indonesia (Olken 2007).

Yet the processes by which long-route accountability may lead to improvements in public services are often conceptualized as involving countervailing power, and programmatic approaches to long-route accountability frequently focus on community members and civic organizations claiming space or working through oversight agencies, courts, or media to shame or punish officials for corruption or other malfeasance.[8] In an influential review of empirical evidence of the effectiveness of social accountability programs, Fox (2015) distinguishes citizen-state collaboration that might enable this kind of countervailing power as involving citizen "voice" augmented by state "teeth."[9] Fox finds that collaborations between public-interest advocacy among citizens and civic organizations and empowered public oversight institutions are able to counter anti-accountability forces more effectively than either citizens or oversight institutions on their own, particularly when these collaborations are large-scale and enduring. He argues that this kind of state-society "sandwich" is particularly important in the context of a "low-accountability trap," where weak horizontal accountability systems lacking checks and balances are further undermined by weak electoral competition that compromises accountability between officials and citizens as their principals.[10]

The second community scorecard program whose effects we explore in this paper, offered in 2017–2018 in 15 communities in Ghana, Malawi, and Sierra Leone, was designed to encourage an alternative kind of state-society collaboration in an alternative context, where political institutions are relatively democratic. Kosack and Fung (2014) hypothesize that political institutions affect the likelihood that transparency and accountability programs will lead to effective long-route state-society collaboration: in contexts where institutions create incentives for long-route accountability, transparency and accountability programs are more likely to lead to long-route collaborative problem-solving between communities and officials to improve public services.[11] Even the most democratically accountable governments contend with principal-agent problems that inhibit optimal delivery of public services responsive to citizens' needs and preferences.[12] When these problems are related to differential power among principals or disproportionate influence of interest groups or organizations (Grossman and Helpman 2001; Moe 1988; North 1990; Olson 1965; Tsebelis 2002), the space offered by an informal, short-term, invited forum encouraged by an external actor may be insufficient to support collaboration between citizens and state actors that leads to improvements (Cornwall 2002, 2004; Fox 2007a, 2015; Gaventa 2006; McGee and Gaventa 2011). But a

range of other issues—including information asymmetries, disconnection and distance between agencies and the citizens they serve, slack, or simple mistrust and misunderstanding between citizens, officials, and providers—are challenges for any bureaucratic organization designed around delegated responsibility and authority, including in democracies (Alchian and Demsetz 1972; Hirschman 1970; Lipsky 1980; Wilson 1991). Both community members and officials may also have free-riding incentives or basic tendencies to ignore or discount capabilities for working together to alleviate these challenges, or simply be unfamiliar with these capacities, particularly in contexts where officials are remote from citizens or where prior state-society relations had been mistrustful.[13] If so, in the context of political institutions designed to create incentives for vertical democratic accountability to citizens, involving officials and leaders in positions of influence and authority in community scorecard programs might broker collaboration between community members and officials that contributes to improvements.

In the large literatures on state-society co-creation and collaborative governance, the implications for effectiveness, efficiency, and responsiveness in government services often stem from state-society embeddedness and mutual support (Evans 1995; Migdal 2001; Ostrom 1990, 1996).[14] One pathway is through social accountability; as noted above, many studies of social accountability argue that requests, incentives, and pressure can lead to citizens, officials, or both more effectively, efficiently, and responsively meeting existing responsibilities. But co-creation may also lead to improvements in public services through innovation, substitution, and self-help: through service providers or other public officials being asked to take on new responsibilities ("innovation"), community members helping with or taking on responsibilities of service providers or other public officials ("substitution"), and community members taking on new responsibilities that were previously neither theirs nor officials' ("self-help") (Creighton et al. 2020). (Table 1.) Any of these four pathways might co-create improvements in public services by relieving information asymmetries, reducing mistrust or slack, or improving efficiency or the allocation and use of resources. In particular, new ideas and deliberation about what is needed to improve a public service, and new efforts to try to realize those improvements, might reduce mistrust or slack by improving incentives, information flows, or oversight, or might clear up misperceptions and improve understanding both among communities of how to use services and among communities and state actors of what is and is not working well, as well as of what each might do to contribute to additional improvements.

Table 1. Co-creation: Four pathways to improved public services

		Who is involved in alleviating the problem?	
		State actors	Communities
Whose responsibility is the problem?	State actors	• Social accountability	• Substitution
	Communities	• Innovation	• Community self-help

To the extent that these collaborations are successful, the process of finding opportunities through discussion and deliberation for citizens, officials, or both to be involved in improving public services, and of learning what each is and is not able to do to improve those services, might also lead each to update their perceptions of the other. Sabel (1993), studying industrial policymaking in Pennsylvania, finds that over several years initially wary industrial firms and state authorities were able to break down barriers of mistrust among themselves and build trust and social capital by demonstrating their willingness not to exploit the others' vulnerabilities. Examining attitudes and approaches toward

environmental stewardship in India, Agrawal (2005) argues that both can change even more quickly, over as little as a year or two, when daily experience gradually leads to updating of existing beliefs and expectations about environmental change and collective efficacy in adjusting to it.

As noted above, there is considerable debate among both scholars and development practitioners about whether the kind of invited forum offered by a community scorecard program can offer space for collaboration of these kinds between citizens and state actors. Many question the efficacy of transparency and accountability programs that are short-term and collaborative and are not complemented with additional resources or with efforts to improve state oversight and other state capabilities. In any context, interactions between citizens and the state that are short-term and informal may offer commensurately fewer opportunities for state actors to, or guarantees that state actors will, take into account citizens' preferences and perspectives.[15] Even in contexts where political institutions are relatively democratic, officials may lack incentives or other reasons to collaborate with citizens to improve public services.[16] If so, a community scorecard program involving facilitated meetings that bring together interested citizens and officials may lead to no more improvement in public services than a community scorecard program without this brokered connection.

3. Two Community Scorecard Programs in Five Countries

In order to examine the hypothesis that involving officials and leaders in positions of influence and authority in a community scorecard program brokers state-society collaboration that improves public services, we compare two community scorecard programs in five countries: one in Indonesia and Tanzania, in which facilitators from local nongovernmental organizations organized meetings in communities, and a second in Ghana, Malawi, and Sierra Leone, in which facilitators also organized several meetings between community members who participated in the community meetings and district-level traditional, elected, and appointed leaders and officials in positions of influence and authority. Ghana, Indonesia, Malawi, Sierra Leone, and Tanzania are a widely diverse group of country contexts in which to explore the observable implications of this hypothesis—economically, politically, culturally, geographically, in the development of their health systems, and in many other ways. Table 2 summarizes several contextual differences across the five countries; see Appendix A for additional discussion.

Table 2. Five country contexts

	Community scorecard (large scale)		Community scorecard with officials (small scale)		
(Initial year of program)	Indonesia (2015)	Tanzania (2015)	Ghana (2017)	Malawi (2017)	Sierra Leone (2017)
Population (millions)	258.2	53.9	28.8	18.6	7.6
GDP per capita (2011 PPP $)	$10,368	$2,491	$4,228	$1,095	$1,390
Democratic constraints on executive authority (–10 to 10 on Polity scale)	9	3	8	6	7
Human Development Index (0 to 1 scale)	0.686	0.528	0.592	0.477	0.419
HDI among women	0.657	0.505	0.563	0.46	0.389
HDI among men	0.709	0.547	0.619	0.492	0.446
EDUCATION					
Primary gross enrollment ratio (% of primary-school-age population)	105	104	145	124	81
Mean years of schooling	7.9	5.8	7.1	4.5	3.5
HEALTH CARE					
Infant mortality (deaths per 1,000 live births)	22.9	41.5	35.7	38.5	81.7
Maternal mortality (deaths per 100,000 live births)	126	398	319	634	1,360
Life expectancy at birth (years)	69	65	63	63.7	52.2
Life expectancy among women	71.2	66.8	64.1	66.2	52.8
Life expectancy among men	67	63.1	62	61	51.6
Current health expenditure (% of GDP)	3.3	3.3	9.6	13.4	6.1

Sources: World Bank (2020), except Polity data, which is from the Polity IV Project (Marshall and Gurr 2014).

Our methodological approach to exploring these observable implications, described in Section 4 below, is rooted in the comparative methods of agreement and difference: we identify differences in the experiences of participants in these two programs that are consistent across diverse contexts. Thus it is important to note that relative to other countries where these programs are common, Ghana, Indonesia, Malawi, Sierra Leone, and Tanzania also share two relative similarities of particular potential relevance to how these programs may improve public services and therefore to the generalizability and scope conditions of our findings. First, health systems in all five, although still uneven in quality and accessibility, have long been the focus of intensive resources and reforms and have been improving steadily for decades. Second, Ghana, Indonesia, Malawi, and Sierra Leone share national political institutions designed to create incentives for vertical democratic accountability.

A. Improving Health Systems

As noted above, evidence of the effectiveness of community scorecard programs like these is mixed. One important reason may be contextual: the existing level of development of public services varies widely among places where these programs are offered. An influential study by Björkman and Svensson (2009) finds that in Uganda in 2004, when the country's infant mortality rate was 72 per 1,000 births, a randomized field experiment of a community scorecard program focused on improving maternal and newborn health care, in which rural community members and clinic staff participated in two rounds of meetings that offered information and facilitated forums for discussing and creating a plan for fixing problems the survey revealed, was sufficient to reduce infant mortality by 33% in just one year, and a follow-up study (Björkman Nyqvist, de Walque, and Svensson 2017) finds sustained long-term improvements in other health outcomes. Seven years later, a study of a similar program offered in randomly selected clinics in Sierra Leone, where infant mortality rates were 104 per 1,000 births—even higher than Uganda's in 2004—found significant effects of a similar magnitude, including a 38% reduction in under-five mortality rates (Christensen et al. 2020). Yet in 2015, when Uganda's infant mortality rate had fallen to 35 per 1,000 births, a larger-scale randomized replication of the 2004 program in Uganda finds no significant differences in infant mortality and only small average differences in other health outcomes (Raffler, Posner, and Parkerson 2019).[17]

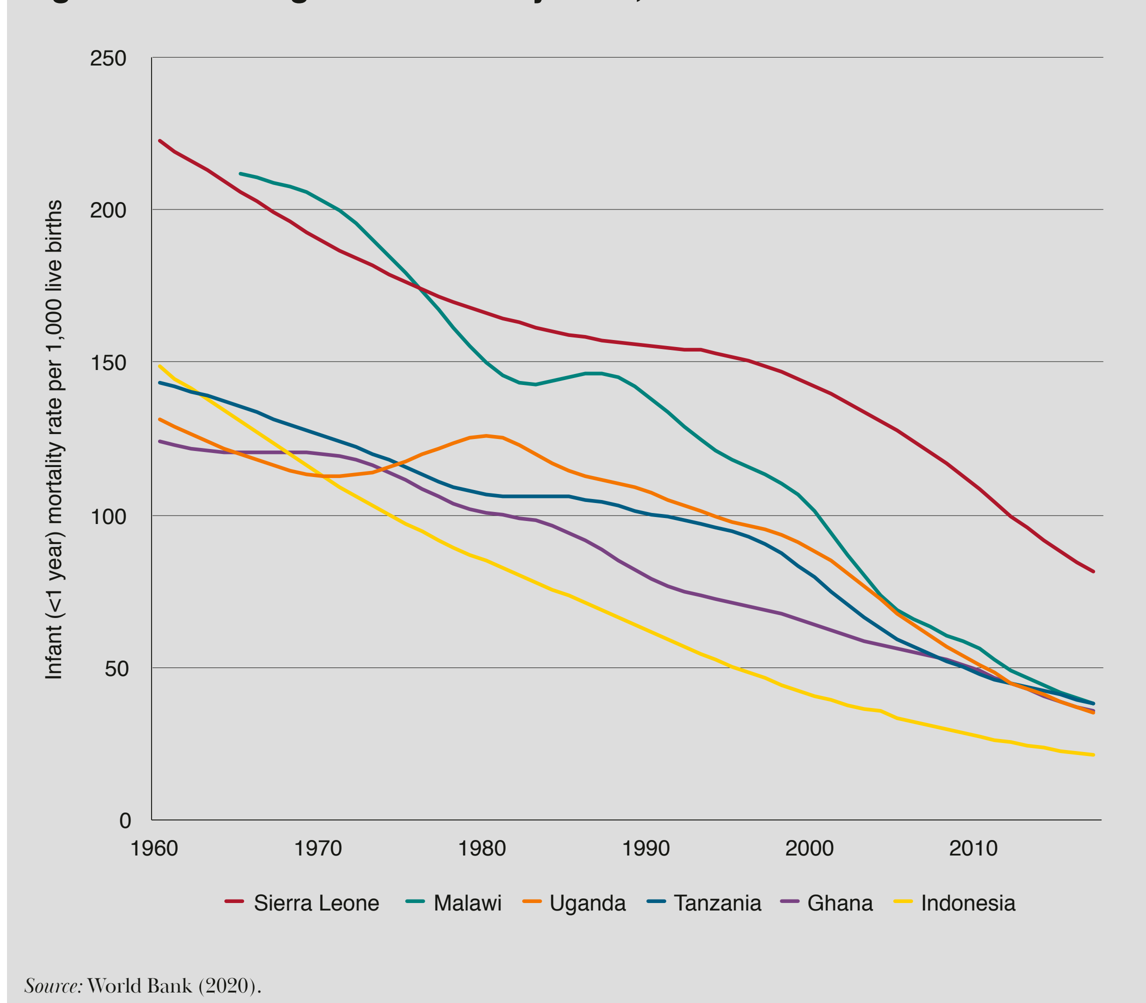
The average impact of the program in Indonesia and Tanzania is consistent with the latter study (Arkedis et al., forthcoming). In almost all the 200 randomly selected communities where the program was offered, community members attended three months of facilitated meetings, discussed the state of maternal and newborn health care, deliberated on what they might do to overcome barriers limiting its quality and access to it, developed plans they could try based on these deliberations, discussed these plans with their broader communities, and reflected on their progress over several months of follow-up discussions. In roughly one-third of communities, participants' efforts led to tangible changes that they and others in their communities noticed at the time and remembered 1.5 years later. For most, participation sustained or increased perceptions of efficacy at improving their communities, and in roughly one-quarter of communities, participants continued their efforts for a year or longer after the facilitator was no longer holding meetings. But on average, the changes their efforts led to were not sufficient for health outcomes in their broader communities to improve significantly further or faster than in a control group of 200 other randomly selected communities in the same regions and provinces where the program was not offered.[18]

In deciding how to try to improve health care, participants in almost all communities focused mostly on what they could do as a community (100% of communities in both countries) and with health-care providers (91% of communities in Indonesia and 74% in Tanzania). They talked with neighbors about the importance of seeking care at clinics or the challenges they faced when doing so; developed clinic transportation pools; or spoke with clinic staff about their accessibility or the way community members were treated at the clinic—among many other approaches. By contrast, participants planned to engage with officials outside their communities in just 29% of communities in Indonesia and 43% in Tanzania. Only 8% of all the specific activities participants planned (6% in Indonesia; 12% in Tanzania) involved officials outside their communities (Creighton et al. 2020).

Yet by the time of the program in Indonesia and Tanzania, health systems in both countries, like the system in Uganda and those in much of the world, had already been the focus of decades of intensive resources and reform efforts.[19] Although still underfunded and of uneven quality and accessibility, both offered lifesaving care and had improved dramatically. Their infant mortality rates at the time—23 and 41 per 1,000 births, respectively—were less like Uganda's in 2004 (72 per 1,000 births) or Sierra Leone's in 2011 (104 per 1,000) than Uganda's in 2015 (35 per 1,000). Figure 1 shows decades of declines

in infant mortality rates in Uganda, Indonesia, and Tanzania as well as Ghana, Malawi, and Sierra Leone; see Appendix A for additional discussion of investments in their health systems.

Figure 1. Declining infant mortality rates, 1960–2018



Source: World Bank (2020).

More developed health systems may have fewer unrealized opportunities for communities using health services to further improve them with short-route efforts. An undeveloped or recently developed health system may have many problems within the capacity of community members to improve, even solve, assuming that they value public health care and do not have an alternative to which to turn to get health care. Community members might, for example, clean the clinic; put up a privacy wall; repair or improve examination rooms, waiting areas, or delivery facilities; or post hours and charges. If those who need medical attention avoid going to a doctor or nurse because they feel disrespected when they do, community members might arrange a conversation with medical staff and ask them how they can work better together. Those who have difficulty reaching the clinic in an emergency might organize a transportation pool or develop a fund that can be used to hire a car. These kinds of solutions might meaningfully improve the quality and accessibility of the care a clinic provides and improve patients'

experiences and their relationships with and trust in providers. Community members' influence might also help with gaps in the knowledge of mothers about more generally accepted tenets of healthy birth practices, such as the importance of delivering in a modern clinic with a trained professional in case something goes wrong, or when and how frequently to seek pre- or postnatal care. Indeed in the 2004 community scorecard program in Uganda, community-focused efforts by participants were found to be *more* effective than those requiring involvement by higher-level officials and other third parties (Björkman Nyqvist, de Walque, and Svensson 2017). Yet after decades of steady improvement, health systems may have fewer of these kinds of issues remaining to solve.

As noted above, communities may also have capabilities for working together with officials above the local level to find solutions to more structural, supply-side problems, such as the facility lacking a steady supply of clean water or electricity, or being too far away to reach in an emergency. If a health clinic keeps short or irregular hours because staff live far away, those who sometimes need emergency or out-of-hours care might offer to welcome staff to homes closer to the clinic or request help from the district health officer or their political representative. In many countries, there are regularly scheduled forums in which community members gather to request development projects for their communities, and an electrical or water supply for their local clinic, or even an entirely new clinic, are among the acceptable requests.

Yet many communities may also discount or be unfamiliar with navigating official channels or forums in which they might request assistance with these kinds of problems. They may be unfamiliar with officials in positions of authority relevant to these problems, be unaware of their authorities and limits, or lack relationships with them or means of approaching them. If so, involving officials in a community scorecard program might offer opportunities for mutually constructive collaboration between citizens and officials that leads to further improvements, even in health systems that have already been the focus of decades of reforms and investments.

B. Competitive Democratic Institutions

The second relative similarity shared by most of the countries we study here is political. At the time of the two programs, Ghana, Indonesia, Malawi, and Sierra Leone shared political institutions that were relatively democratic. None were fully democratic, and in all four the institutions shaping these incentives were also recent: each is a postcolonial state where prior state-society relations were not always collaborative and trusting and where government was still seen by many as relatively remote from citizens' needs. (See Appendix A.) Yet at the time of these two programs, all four had shared more recent experiences with regularly contested democratic elections and multiple peaceful transfers of national political power between elected representatives of rival constituencies with competing political preferences, one of the mechanisms by which democratic institutions are designed to create responsiveness and accountability to citizens.[20] Among the five countries we study, the only one without similar experiences was Tanzania: although it was democratizing rapidly in the years before the program, at the time it was still a "competitive authoritarian" system (Levitsky and Way 2010) in which one party had held power for more than half a century, and in the years after the program it has again become more authoritarian.[21]

With the program involving officials in Ghana, Malawi, and Sierra Leone, this contextual similarity in their political institutions was purposeful. Section 2 noted that transparency and accountability programs may be more likely to encourage long-route collaboration to improve public services in contexts with vertical institutional incentives for responsiveness to citizens than in contexts of low-accountability traps. In addition to the incentives of their national political institutions, in Ghana and Sierra Leone the districts where the program was offered may have had additional institutional incentives for responsiveness. In one, a recent district election had led to many new officials in office, including a majority of those who

became involved in the program. In the other, the district itself was new, created by the president several years earlier in response to pressure from those who lived there for autonomy and development.

In short, these five countries shared health systems that were uneven in quality and accessibility but had all been the focus of decades of intensive resources and reforms and had been improving for decades. Four also shared political institutions that, although not fully democratic, were designed to create incentives for responsive public services and had seen multiple transfers of political power between competing constituencies.

Figure 2 summarizes these two contextual dimensions and their potential influence on differences in the experiences of participants in the two community scorecard programs.

Figure 2. Hypothesis in context

			Similarities (potential scope conditions)	
		Development of the health system	*Prior investments in the health system*	*Political institutions*
Community scorecard (large scale)	Indonesia	Medium-High	Improving for decades	Democratic
	Tanzania	Medium-Low	Improving for decades	Competitive authoritarian but democratizing
Community scorecard with officials (small scale)	Ghana	Medium-Low	Improving for decades	Democratic
	Malawi	Medium-Low	Improving for decades	Democratic
	Sierra Leone	Low	Improving for decades	Democratic

Notes: The solid line represents hypothesized differences between the experiences of participants in the two community scorecard programs. Differences in democratic political institutions and the relative development of health systems (the dotted lines) may also have influenced the experiences of participants similarly across diverse contexts.

C. Two Community Scorecard Programs

The community scorecard programs we study were designed by nongovernmental organizations with experience in transparency and accountability, governance, community engagement, and community health—in Ghana, the Center for Democratic Development (CDD-Ghana); in Indonesia, PATTIRO, a research and policy advocacy organization focused on regional and local governance issues in a number of sectors; the Malawi Economic Justice Network (MEJN); the Water, Sanitation and Hygiene Network (WASH-Net) in Sierra Leone; and the Tanzania country office of the Clinton Health Access Initiative

(CHAI)—as well as by an international team of practitioners and scholars, several authors of this paper among them.[22] Facilitators were not from communities where the programs were offered, but they typically were from nearby communities and spoke local languages. In each community, the facilitator first spent time surveying health facilities and women who had recently given birth and gathering stories of experiences with maternal and newborn health care.[23] The facilitator then invited community members to attend six community meetings held over a period of approximately three to four months.[24] In the first meeting, the facilitator led a discussion of the results of the survey and presented statistics about the care available in the community relative to others in the form of a "scorecard." While the scorecard was presented in a different specific format in each country, the overall structure of the scorecard in each community included a brief description of maternal and newborn health-care-seeking among the community (such as delivering in a facility) and of potential problems with maternal and newborn health care (such as the facility being difficult to get to, a shortage of delivery beds, or the prevalence of misunderstandings among community members about effective care-seeking) and a statistic or graphical representation of care-seeking or the degree of each problem in the community based on the survey the facilitator had conducted. In addition to sharing the scorecard, the facilitators told several stories of how other communities around them had tried to improve their public services. After this discussion, facilitators were trained to ask those still interested to come back for a second meeting in which they deliberated on a plan of activities that might improve those services or increase access to or awareness and use of them among members of their community. See Appendix B for examples of scorecards and Appendix C for examples of participants' plans. Before departing the community, facilitators also organized a third meeting in which participants in the first two meetings presented and discussed their plan with those in their broader community who were interested in learning about it or becoming involved. After departing the community, facilitators returned for three more discussions, held approximately one, two, and three months later, during each of which they encouraged anyone still involved to reflect together on how their plan had worked, make adjustments, and plan for how to maintain their efforts once the facilitator was no longer organizing meetings. Although the two programs were otherwise designed and iteratively piloted to be appropriate for each country, region, and district, these basic elements were largely the same in both programs.[25]

In the program in Ghana, Malawi, and Sierra Leone, facilitators and other staff members of the nongovernmental organizations organized and helped to facilitate two additional meetings between several participants from each community and officials and leaders in positions of influence and authority who had expressed a willingness to be involved in helping them improve maternal and newborn health care in their communities.[26] The meetings had four purposes: 1) to provide opportunities for community participants and officials to meet; 2) to offer opportunities for community participants to learn more about health-care governance, including the organization of the health system and of functions, rights, and responsibilities within it; 3) to provide venues for community participants to share their plans, describe their progress, and receive feedback; and 4) to offer forums for community participants to ask officials for assistance with their efforts. To prepare community members for these meetings, facilitators in some communities also spoke about policies in the health system, rights of citizens to health care, and the roles of government officials.[27] The first of the two additional meetings with officials was midway through the program, after participants had developed and begun to try their plans for improving their community's health care. The second was held at the end of the program, close to the final meeting. The organizations organizing these meetings had not worked previously in 13 of the 15 communities where the program was offered.[28] But all three were well-known national organizations with reputations and relationships that facilitated their brokering of these meetings with officials; in Ghana and Sierra Leone, the organizations had existing relationships from prior work in the district with several of the officials who became involved.

The program involving officials took two distinct approaches to brokering collaboration between citizens and officials, a difference we find below to have been particularly important to our conclusions about differences in the experiences of participants.[29] In Ghana and Sierra Leone, the officials involved were elected and appointed officers with formal responsibility for and authority over district public services. In Ghana, the officials involved included the District Director of Health Services, who has direct authority over health facilities, as well as the Presiding Member for the District Assembly and the District Planning Officer. In Sierra Leone, the officials involved included the District Development and Planning Officer, the District Chief Administrator, and three Local Councillors who represented the five communities who were offered the program.[30] (See Appendix B.)

Rather than elected or appointed officials, the program in Malawi brokered collaboration primarily with traditional authorities. Ghana, Malawi, and Sierra Leone all have structures of hereditary traditional authority parallel to other government structures.[31] In Malawi, traditional authorities have customary responsibilities around land allocation and adjudication of disputes, often interface with nongovernmental organizations and help them distribute aid, and have influence in community health and relationships with government officials, including those with formal authority over the health system. But traditional authorities do not control public funds or taxes; rather their influence over day-to-day functions of district government, including the health system, is mostly informal, indirect, and derived from respect for their positions within communities (Walsh et al. 2018). In three communities in Malawi, the community scorecard program involved meetings with the areas' Traditional Authorities—relatively high-level traditional authorities who also sit on the District Executive Committee with officials who have authority over district health and other public services—and in two, Group Village Heads—traditional authorities who lead multiple villages and are responsible to the area's Traditional Authority (see Appendix B).[32] The second interface meeting in Malawi, at the end of the program, was incorporated into a regular meeting of the District Executive Committee with other district-level officials. But until this meeting at the end of the program, the only officials in Malawi with whom facilitators organized meetings were traditional authorities.

Both community scorecard programs were offered in mostly rural communities selected randomly from groups of communities stratified similarly, such that their access to maternal and newborn health care was not systematically different from others in the region, province, or district. In the earlier program in Indonesia and Tanzania, one or two communities were randomly selected from the catchment areas of facilities across two regions or provinces in each country. In the smaller program in Ghana, Malawi, and Sierra Leone, the five communities in each country who were offered the program were randomly selected from the catchment areas of five health facilities in a district (administrative areas smaller than the regions in Tanzania and provinces in Indonesia). In all five countries, a second control group of communities—100 communities each in Indonesia and Tanzania, 3 each in Ghana, Malawi, and Sierra Leone—were also randomly selected where health-care outcomes were measured but the community scorecard programs were not offered. The scale of the scorecard program in Ghana, Malawi, and Sierra Leone is too small (five communities in each) to draw internally valid inferences about whether the program had more significant average effects on community-level health outcomes than the program in Indonesia and Tanzania, but the small control groups in Ghana, Malawi, and Sierra Leone allow us to ask whether any observed changes to health care were encompassing or specific to communities where the program was offered. Appendix A describes the process of random selection.

4. Methods

In order to explore the hypothesis that the program involving officials brokered collaboration that led to more improvement in maternal and newborn health care, we examine the extent to which differences in participants' experiences of these two community scorecard programs were consistent with a range of observable implications of this hypothesis. The difference in scale between the two programs—200 communities in two countries for the first; 15 communities in three countries for the second—makes average comparisons unreliable. Instead we follow a methodological approach based in the comparative methods of agreement and difference (Mill 1843): we ask, through comparative process tracing, whether differences in participants' experiences in these two (large or small) groups of communities across diverse countries are consistent or inconsistent with observable implications of this hypothesis.[33] In particular, we focus on five questions about participants' experiences in these two programs:

1. Across diverse communities, did community members participate meaningfully in the program—discussing and deliberating on the information the facilitator provided, and planning and attempting to improve public services or convince others to—and find the experience helpful for improving their community's maternal and newborn health care?
2. Did participants in the program that involved officials work more extensively with officials?
3. Did officials involved respond constructively to participants' attempts to work with them?
4. Did more collaboration between community participants and officials contribute to more changes in maternal and newborn health care?
5. To what extent do these answers generalize across contexts?

The observable implications of the hypothesis for the five questions posed above across these contexts are summarized in Table 3.

Table 3. Observable implications of hypothesis across contexts

	Community scorecard (large scale)	Community scorecard with officials (small scale)
Across diverse communities did community members participate meaningfully in the program and find the experience helpful for improving their community's health care?	Yes	Yes
Did participants in the program that involved officials work more extensively with officials?	Less extensively	More extensively
Did officials respond constructively?	N/A	Yes
Did more collaboration between community participants and officials contribute to more changes in maternal and newborn health care?	N/A	Yes
To what extent do these answers generalize across contexts?	Across diverse contexts, including health-care system development and political institutions. Potential scope conditions: long-improving health-care systems neither among world's most nor least developed.	Across diverse contexts, including health-care system development. Potential scope conditions: long-improving health-care systems and democratic political institutions.

Any systematic perspective on these observable implications inevitably has observer, social desirability, and other biases in reliably reflecting how participants in diverse places experienced these programs. Thus we focus less on any particular comparison or metric than on patterns in the experiences of participants that are consistent from multiple perspectives. In particular, we integrate six perspectives from interviews, focus groups, and observations, each systematically replicable across these diverse contexts and each offering distinct perspectives on the observable implications in Table 3:[34]

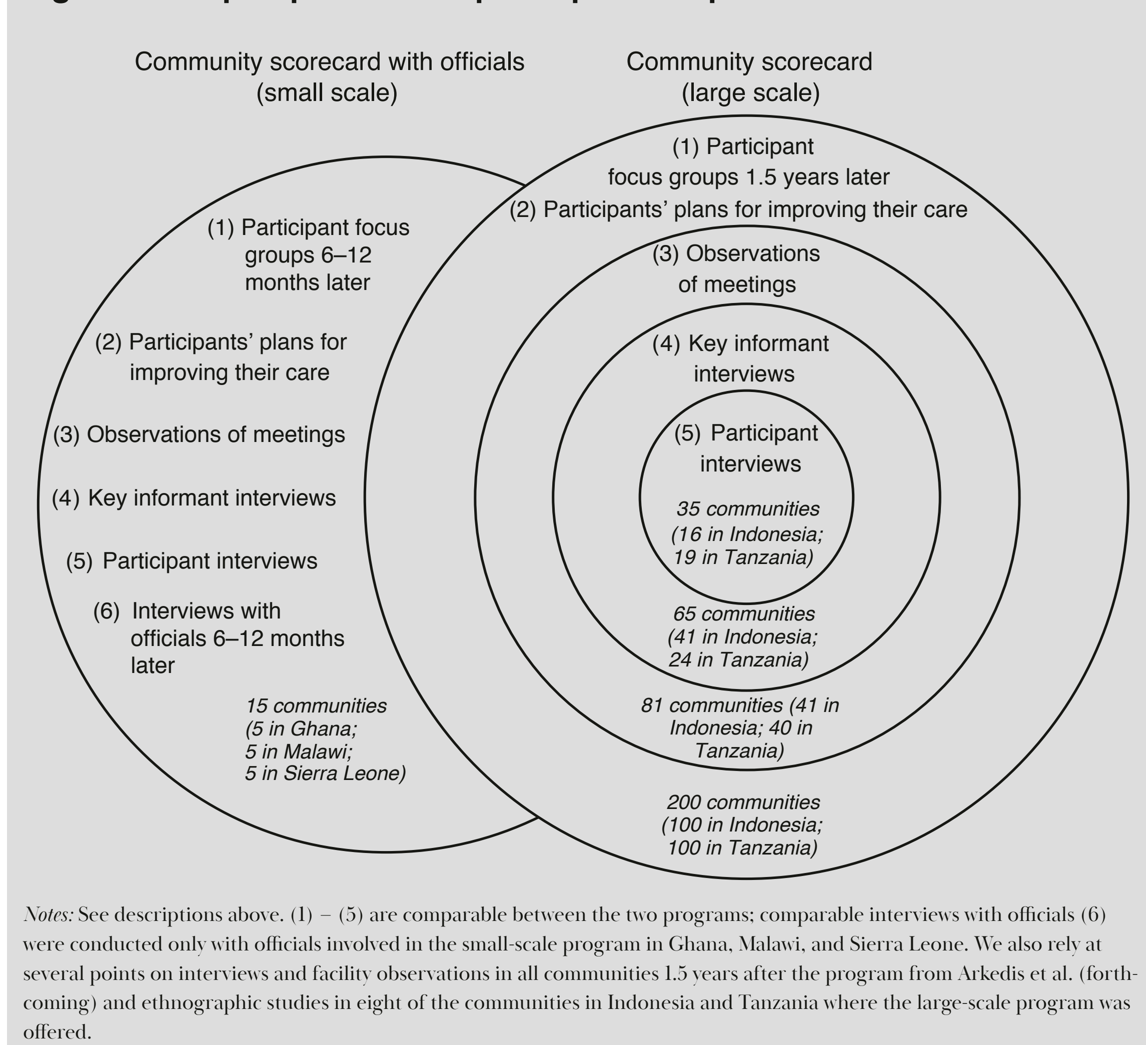
1. Six months after the program in Sierra Leone, a year after the program in Ghana and Malawi, and 1.5 years after the program in Indonesia and Tanzania, we invited participants in all communities to a focus group in which we asked them to reflect on the program and any activities they remembered, including challenges they had faced and any changes they had seen as a result of the program. We also asked whether participants thought that their efforts had improved their communities' health care overall, whether they were still trying to improve their care or their communities in other ways, whether they had experienced any personal benefits or costs from participating, and overall whether or not they were glad that they had participated.
2. We analyzed plans of specific activities for improving health care developed by participants in the meetings in each community.[35]

3. In 81 of the 200 communities in Indonesia and Tanzania and all 15 communities in Ghana, Malawi, and Sierra Leone where these two programs were offered, the authors and other trained observers attended several of the meetings and answered a series of questions about engagement, discussion, and decision-making in them: both interface meetings; the first and second community meetings, at which participants discussed the information provided and deliberated on activities to pursue; and the third and final follow-up meeting in each community, at which participants discussed their progress over the previous three months and made plans to sustain their activities after the facilitator had left.
4. In all 15 communities in Ghana, Malawi, and Sierra Leone, in 41 communities in Indonesia, and in 24 communities in Tanzania, we or trained observers asked follow-up questions after the final program meeting of several participants and of health providers, neighbors, officials, or anyone else with whom they planned to engage as part of their activities, and observed any effects that they described.
5. In all 15 communities in Ghana, Malawi, and Sierra Leone as well as 35 of the 200 in Indonesia and Tanzania (16 in Indonesia and 19 in Tanzania), we interviewed all participants before the first meeting and after the last about expectations for and reflections on their participation and the efficacy of their efforts. In Ghana, Malawi, and Sierra Leone, participants were also asked to identify which of four anchoring vignettes (King et al. 2004), in which a small group of people in a community like theirs tried to work with officials to improve their school, was most familiar to them.[36]
6. Six to twelve months after the program in Ghana, Malawi, and Sierra Leone, we asked all officials and leaders who had been involved to reflect on the program, any activities they remembered, challenges they had faced, changes they had seen, and the costs and benefits of the program for them and for improving health care in communities in their jurisdictions. To avoid biasing answers and to leave space for officials and leaders to share honest reflections, interviews involved only a few open-ended questions, without specific prompts to recall specific activities, challenges, or changes.[37]

Figure 3 summarizes these six perspectives on participants' experiences of the two programs. Integrating them offers multiple perspectives on the observable implications in Table 3, such that each partly compensates for biases in the others in reliably reflecting how participants in each country responded to and experienced these programs. Comparisons of observations of meetings in the communities offered each program allow us to ask whether those attending engaged in substantive discussion and deliberations, as well as whether they tried any of the approaches they had planned to improve maternal and newborn health care, and can provide suggestive evidence of their skepticism or optimism about whether their efforts could and were improving their maternal and newborn health care. Comparisons of interviews with participants in the two programs offer suggestive evidence of participants' perceptions of their own civic efficacy as well as of the responsiveness of officials. Stratified random selection of communities in each country limited contextual bias in access to maternal and newborn health care in these communities relative to other communities in the province, region, or district.[38] To reduce observer bias, meetings were observed and interviews and focus groups conducted by authors of this paper based in each country. The exception is the final interviews 6–12 months after the program with officials who were involved, which were conducted by an author from the United States. Comparisons of interviews with participants and others with whom they tried to engage, along with focus groups with participants between six months and 1.5 years after the two programs, help to further check against observer biases by verifying that the activities participants described in meetings actually

occurred as well as similarities with participants' perceptions of whether they had led to any changes in their community's health care.

Figure 3. Six perspectives on participants' experiences



Notes: See descriptions above. (1) – (5) are comparable between the two programs; comparable interviews with officials (6) were conducted only with officials involved in the small-scale program in Ghana, Malawi, and Sierra Leone. We also rely at several points on interviews and facility observations in all communities 1.5 years after the program from Arkedis et al. (forthcoming) and ethnographic studies in eight of the communities in Indonesia and Tanzania where the large-scale program was offered.

In short, in five countries a group of five or one hundred communities were randomly selected from among mostly rural communities stratified to be representative of access to maternal and newborn health care in each district, province, or region where the program was offered. The five countries in which the two programs were offered differ on myriad dimensions, including in health-system outcomes. Yet they also share contextual similarities of potential importance to our inquiry into participants' experiences of these two programs. All had health systems that, although still uneven in quality and accessibility, had been the focus of decades of intensive resources and reforms and had been improving for decades. In addition, four shared democratic political institutions; the fifth was a competitive authoritarian regime that at the time was democratizing.

5. Findings

The experiences of participants in the two programs were similar in several important respects (Table 4). In both programs, participants in almost every community created plans aimed at a wide range of supply-side and, other than in Sierra Leone (Table 6 below), demand-side improvements that involved a range of activities (an average of 3 distinct activities in Malawi, 4 in Tanzania, 6 in Ghana and Sierra Leone, and 7 in Indonesia), and in almost every community participants tried at least some of what they had planned.[39] By the final meeting, participants in just over half of communities in Tanzania seemed to observers to have become discouraged and skeptical. But in the other half of communities in Tanzania, and in the vast majority of communities in the other four countries, participants appeared optimistic in the final meeting that their efforts would sustain improvements in their community's health care. Reflecting on their experiences in focus-group discussions between six months and 1.5 years after the program, in many communities most or all (>75%) of the individual participants recalled at least one of their efforts as unsuccessful: 39% of communities in Indonesia, 41% in Tanzania, 60% in Ghana and Malawi, and 100% in Sierra Leone. But participants in almost all communities in all five countries also discussed specific efforts in these focus groups that most (>75%) recalled as successful at meeting their goals (93% of communities in Indonesia, 100% in Tanzania, 60% in Sierra Leone, 80% in Ghana, and 100% in Malawi), and participants as a group generally recalled their efforts as having improved health care in almost all communities (83% in Indonesia, 95% in Tanzania, and 100% in Ghana, Malawi, and Sierra Leone). In a substantial proportion of communities, participants in these focus groups also described continuing their efforts months or years after the facilitator was no longer holding meetings (23% in Indonesia, 26% in Tanzania, 40% in Malawi, 60% in Ghana, and 100% in Sierra Leone). Participants recalled their participation as personally beneficial in 80% of communities in Malawi, 97% in Indonesia, and 100% in Tanzania, Ghana, and Sierra Leone. In short, with several caveats we discuss below, the answer to our first question—did community members living in diverse communities participate meaningfully in the program and find the experience helpful for improving their community's health care—is yes for both programs.

Table 4. Planning, perceptions of efficacy, and later reflections on participating

	Community scorecard (large scale)		Community scorecard with officials (small scale)		
	Indonesia	Tanzania	Ghana	Sierra Leone	Malawi
Participants who, prior to the program, described themselves as fully or mostly capable of making improvements in their communities (% of participants)*	39%	98%	100%	94%	100%
Average number of distinct activities in participants' plans	7	4	6	6	3
Communities in which, at the 3rd follow-up meeting, most or all participants appeared optimistic to observers that their efforts would sustain improvements in their community's health care** (% of communities)	88%	48%	100%	100%	80%
Communities in which, in focus groups after the program . . . (% of communities)	(1.5 years later)	(1.5 years later)	(1 year later)	(6 months later)	(1 year later)
. . . >75% of individual participants recalled specific efforts that were successful at meeting their goals	93%	100%	80%	60%	100%
. . . >75% of individual participants recalled specific efforts that were unsuccessful at meeting their goals	39%	41%	60%	100%	60%
. . . participants as a group recalled their efforts as having generally improved health care in their communities	83%	95%	100%	100%	100%
. . . participants as a group said that they were . . .					
. . . continuing to meet	23%	26%	60%	100%	40%
. . . glad they participated	97%	100%	100%	100%	80%

Notes: See Appendix D for interview and focus-group questions as well as further evidence from interviews, focus groups, and meeting observations of participants' engagement in meetings, planning, confidence in efficacy, and reflections on participating. * Data based on 16 communities in Indonesia and 19 in Tanzania. ** Data based on 41 communities in Indonesia and 40 in Tanzania.

The answers to the other questions in Table 3 are more complicated. Section A below describes a range of evidence suggesting that in Ghana and Sierra Leone, the addition of meetings with officials in positions of influence and authority who had expressed interest in participating led to differences that were generally consistent with the other observable implications of the hypothesis. Section B describes comparable evidence that these additional meetings did not lead to similar differences in Malawi.

A. Ghana and Sierra Leone: A Community Scorecard Program Involving District Officials

In the program in Ghana and Sierra Leone, the addition of meetings with officials in positions of influence and authority who had expressed interest in participating led to several consistent differences in the experiences of community participants from the experiences of participants in Indonesia and Tanzania.

Working with officials

Participants in more communities worked more often with government officials in Ghana and Sierra Leone than in the program in Indonesia and Tanzania. In every community in Ghana and Sierra Leone, participants' plans included steps to engage with officials, and more than half of all the distinct activities that participants in Ghana and Sierra Leone planned involved officials outside their communities; in the program in Indonesia and Tanzania, less than 10% of what participants planned involved officials outside their communities (see Tables C.3–C.6 in Appendix C). Most officials were the district officials with whom participants connected during the two interface meetings the facilitator organized. But not all: in one of the five communities in Ghana and two in Sierra Leone, participants also tried working with their Members of Parliament, after the MPs expressed interest in being involved in the meetings. In three of the five communities in Ghana and four of the five in Sierra Leone, interviews and observations confirm that participants did engage with officials. Proportionately this is not substantially higher than in the program in Tanzania—where in 43% of communities participants planned to involve officials in at least one of their activities, and 1.5 years later participants in 53% recalled attempting to engage with officials—but is substantially higher than in Indonesia, where participants in 29% of communities planned at least one activity involving officials, and 1.5 years after the program participants in 35% recalled at least one attempt to engage with officials.

What did these interactions involve? Table 5 classifies participants' plans by the typology in Table 1 above: activities that sought improvements through "social accountability," involving service providers or officials being asked to do something because it is their responsibility; "innovation," involving service providers or officials being asked to take on a new responsibility; "substitution," involving community participants or others in their communities taking on activities that are the responsibility of service providers or officials; and "community self-help," involving participants or others in their community taking on new responsibilities that were previously neither theirs nor providers' or officials'.

Table 5. Planned activities to improve health care

	Community scorecard (large scale)		Community scorecard with officials (small scale)		
	Indonesia	Tanzania	Ghana	Sierra Leone	Malawi
Social accountability	29%	20%	34%	55%	40%
Innovation	24%	33%	47%	36%	53%
Substitution	4%	6%	9%	18%	0%
Community self-help	55%	46%	22%	24%	47%

Notes: Figures are proportions of planned activities. Proportions do not sum to 100% because classifications are not mutually exclusive and activities could have multiple objectives in different categories. (Multiple objectives were more common in the community scorecard program with officials.)

Other than activities that sought to substitute for the responsibilities of others, which were rare, participants' plans in all five countries involved new or expanded attempts to co-create improvements with officials or work with officials in cooperative governance through social accountability, innovation, and community self-help (Evans 1995; Fox 2015; Migdal 2001; Ostrom 1990, 1996). But in Ghana and Sierra Leone, participants' plans were relatively more likely than in Indonesia and Tanzania to seek improvements through social accountability and innovation and relatively less likely to rely only on community self-help. Participants' plans were relatively more likely to involve social accountability for existing responsibilities in Sierra Leone (55% of planned activities) than in Indonesia and Tanzania (29% of planned activities in Indonesia and 20% in Tanzania). (See Table 5 above.) But in all five countries, most of participants' planned activities—including planned efforts at social accountability—also relied on participants, providers and/or officials, or both taking on new efforts or responsibilities: 79% of activities in Indonesia; 82% in Tanzania; and 75% in Ghana. Even in Sierra Leone, where participants were relatively more likely to rely on social accountability than participants in the other countries, 52% of participants' planned activities also involved new community, provider, or official efforts or responsibilities.

Compared with participants in Indonesia and Tanzania, participants in Ghana and Sierra Leone also planned more activities involving officials in order to realize almost all of their objectives for improving their maternal and newborn care. Table 6 shows proportions of communities where participants planned activities with different objectives involving officials.[40]

Table 6. Objectives of participants' plans

	Community scorecard (large scale)		Community scorecard with officials (small scale)		
	Indonesia	Tanzania	Ghana	Sierra Leone	Malawi
SUPPLY SIDE	98% (30%)	92% (49%)	100% (100%)	100% (100%)	100% (40%)
Health facility	95% (29%)	79% (47%)	100% (100%)	100% (100%)	80% (20%)
Facility infrastructure	32% (11%)	24% (7%)	40% (40%)	80% (80%)	20% (0%)
Access (e.g., new roads; mobile or outreach services; longer hours; new ambulance; community transportation pools)	69% (19%)	32% (5%)	80% (60%)	80% (80%)	20% (0%)
New facility	26% (2%)	43% (19%)	20% (20%)	0% (0%)	0% (0%)
Availability of drugs, supplies and other inputs	45% (8%)	11% (8%)	80% (80%)	100% (80%)	60% (20%)
Facility staffing	16% (2%)	19% (15%)	60% (60%)	20% (20%)	0% (0%)
Provider knowledge	2% (0%)	0% (0%)	0% (0%)	0% (0%)	0% (0%)
Patients' experiences	68% (6%)	52% (7%)	60% (40%)	60% (20%)	100% (40%)
Availability, attitude, and effort of or trust in providers	41% (2%)	31% (4%)	60% (40%)	40% (20%)	100% (40%)

	Community scorecard (large scale)		Community scorecard with officials (small scale)		
	Indonesia	Tanzania	Ghana	Sierra Leone	Malawi
Information (e.g., cost; opening hours) or complaint mechanisms	42% (3%)	36% (3%)	0% (0%)	0% (0%)	0% (0%)
Facility cleanliness	10% (1%)	2% (0%)	0% (0%)	20% (0%)	0% (0%)
DEMAND SIDE	96% (4%)	99% (5%)	100% (80%)	40% (0%)	100% (40%)
Awareness, knowledge, and attitudes	91% (0%)	95% (0%)	60% (20%)	20% (0%)	100% (20%)
Ability to pay	44% (4%)	46% (1%)	60% (60%)	40% (0%)	20% (0%)
Bylaws, partnerships, or other efforts to increase use of care	16% (0%)	54% (4%)	0% (0%)	0% (0%)	20% (20%)
Non-health-system directed	18% (0%)	0% (0%)	0% (0%)	0% (0%)	0% (0%)
Not health related	0% (0%)	0% (0%)	40% (20%)	40% (40%)	0% (0%)

Note: Proportions in parentheses are communities where participants' plans included them attempting to involve officials above the community level or asking for help from others in involving officials above the community level.

Table 6 shows consistent differences between the programs in whether participants' plans for realizing these objectives involved officials above the community level. In both programs, participants' plans included supply-side improvements to maternal and newborn health care (98% of communities in Indonesia, 92% in Tanzania, and every community in Ghana and Sierra Leone). But participants in Indonesia and Tanzania planned to involve officials in these efforts in less than half of communities (30% of communities in Indonesia and 49% in Tanzania); in Ghana and Sierra Leone, by contrast, participants in every community planned to involve higher-level officials in realizing supply-side improvements. In Ghana, participants were also substantially more likely to involve officials in demand-side efforts: in 80% of communities in Ghana, participants' demand-side plans involved officials, compared with 4% in Indonesia, 5% in Tanzania, and neither of the two communities in Sierra Leone where participants had demand-side objectives.[41]

Participants in Ghana and Sierra Leone also focused more than participants in Indonesia and Tanzania on aspects of care related to supply chains, facility staffing, and facility infrastructure, with which officials might be helpful or over which they had formal authority. In both programs, participants' plans almost always included efforts to improve health facilities (95% of communities in Indonesia, 79% in Tanzania, and every community in Ghana and Sierra Leone). But in Ghana, participants in more communities focused on facility staffing than participants in any other country (60% of communities in Ghana compared with 20% in Sierra Leone, 16% in Indonesia, and 19% in Tanzania); in Sierra Leone, participants were more likely than in any other country to focus on facility infrastructure (80% of communities in Sierra Leone compared with 40% in Ghana, 32% in Indonesia, and 24% in Tanzania); and in both Ghana and Sierra Leone, participants in most communities also focused on supply chains for drugs and supplies (80% of communities in Ghana and every community in Sierra Leone—compared with 45% of communities in Indonesia and 11% in Tanzania). The only objective related to improving health facilities that was more common in Indonesia and Tanzania than in Ghana and Sierra Leone was

to seek an entirely new health facility: only one community in Ghana and none in Sierra Leone planned to seek a new facility, compared with 26% of communities in Indonesia and 43% in Tanzania.

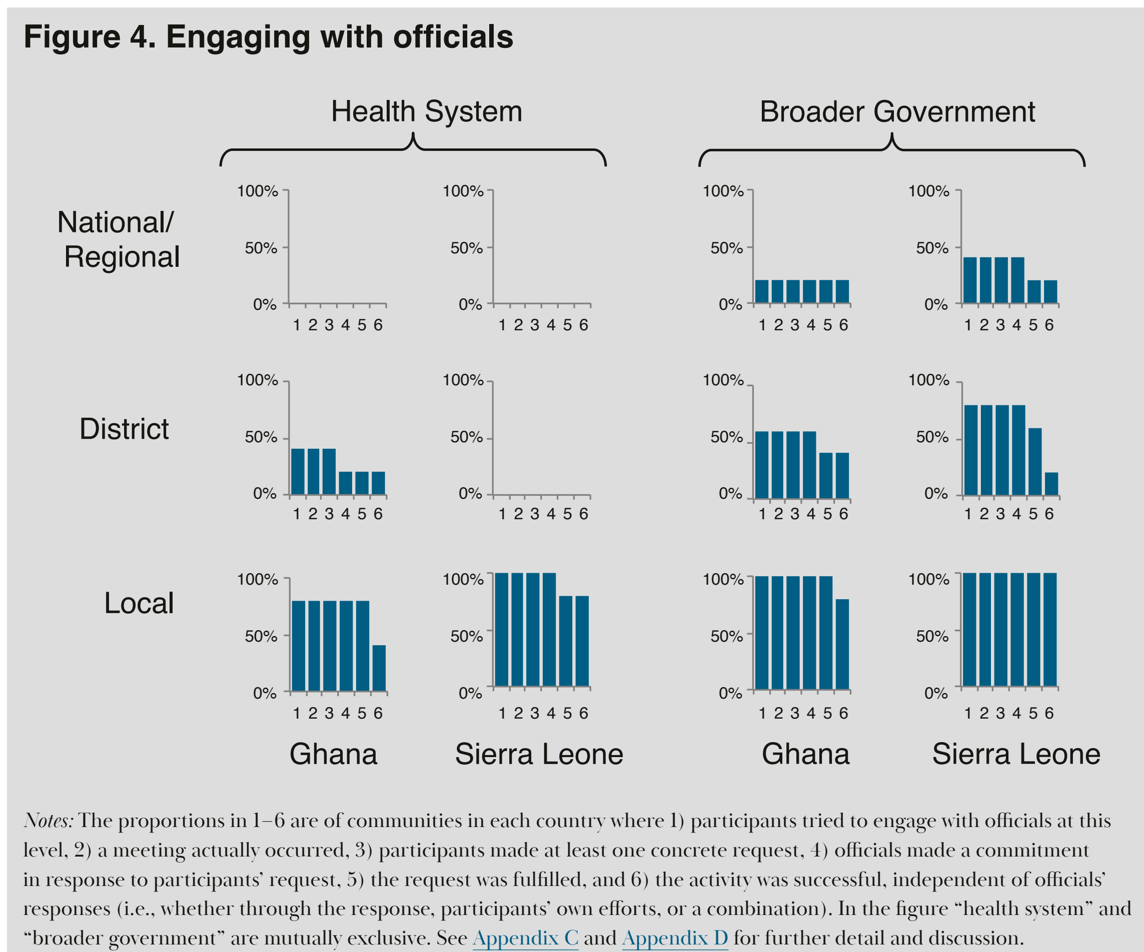
In short, the answer to our second question—whether participants in the program that involved officials worked more extensively with officials—is yes for the program in Ghana and Sierra Leone.

Officials' responses

In both Ghana and Sierra Leone, the two meetings that facilitators organized between officials and community members involved 30–50 people, lasted two to three hours or more, and included at least two participants from each community and most although not all of the officials who at the start of the program had expressed interest in the program and willingness to be involved.[42] Observations of the meetings and interviews with participants suggest that the discussions involved participants from each community sharing plans and discussing progress, and officials providing advice and suggestions on what participants could do to make further progress, clarifying policies or practices in the health system, and discussing their roles, responsibilities, and capacities,[43] and that several participants asked officials for help with their plans.[44]

Officials rarely made promises in the interface meetings. But interviews and observations generally suggest that officials in both Ghana and Sierra Leone eventually responded constructively to community participants' requests. Figure 4 summarizes these responses by showing the proportions of communities in which 1) participants tried to engage with officials, and in which interviews and observations indicate 2) that officials met with participants at program meetings or outside them; 3) that participants made concrete requests; 4) that officials responded with specific commitments to these requests; 5) that officials met these commitments; and 6) that, independent of this response (i.e., regardless of whether results were owing to the officials' responses, participants' own efforts, or both), participants saw their goals realized—for example, new beds, ambulances, staff, rooms, or water supplies; improved roads; improved relations with providers; or changes in other areas or outside the health system. In both countries, in every community where participants tried to work with officials, they made concrete requests for help. In all but one, officials committed to responding, and in all but three the officials fulfilled at least some of these commitments.

Figure 4. Engaging with officials



Notes: The proportions in 1–6 are of communities in each country where 1) participants tried to engage with officials at this level, 2) a meeting actually occurred, 3) participants made at least one concrete request, 4) officials made a commitment in response to participants' request, 5) the request was fulfilled, and 6) the activity was successful, independent of officials' responses (i.e., whether through the response, participants' own efforts, or a combination). In the figure "health system" and "broader government" are mutually exclusive. See Appendix C and Appendix D for further detail and discussion.

Six months (in Sierra Leone) or twelve months (in Ghana) after the program, we also interviewed all officials who had stayed involved through the final meeting with open-ended questions about their experiences. In these interviews, officials recalled nine specific responses to participants' efforts in Ghana and six in Sierra Leone. In Ghana, by triangulating with other interviews and observations, we were able to strongly verify three of these specific responses and weakly verify a fourth.[45] One official worked with the District Assembly to open a new facility that had sat unopened for three years because beds had not been delivered. Another arranged for representatives of the National Insurance Scheme to explain fees and charges and used supervisory authority to repair relations between facility staff and community members concerned about fees they were paying. This official also worked with participants in two communities in conflict over the site of their facility on a new facility up to government standards that would be located between the two. Following their meeting, participants worked with the elders in their community to resolve the feud and to identify and clear land for the facility, and worked with the official to request and secure funding for the facility. We later observed the new facility to be under construction.

In Sierra Leone, officials described six specific responses, of which we were able to strongly verify two and have some triangulating evidence of a third. Unlike officials' responses in Ghana, none relied on government resources—a difference consistent with Sierra Leone's relatively less developed economy and health system. Instead officials in Sierra Leone focused on making connections and resolving conflict. One official connected participants with a grant writer in the district, who helped them secure

assistance from a nongovernmental organization to repair the pump at their facility and dig a second well to improve the facility's supply of water. A second official helped mediate a conflict between a community and facility staff that led to a substantial, although short-term, improvement in their relationship and their perceptions of how they were treated at the facility. (See Appendix E for more detailed descriptions of each.)

In short, the answer to the third question—whether officials involved responded constructively to participants' attempts to work with them—is yes for the program in Ghana and Sierra Leone.

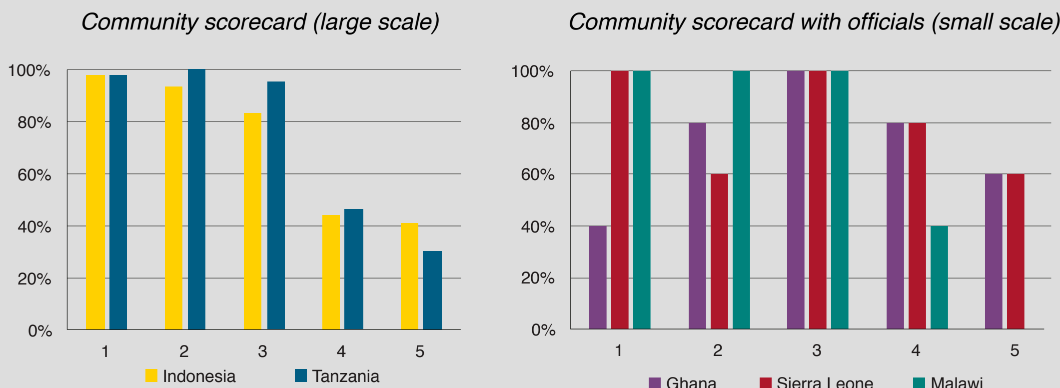
Collaboration and change in maternal and newborn health care

These responses were sufficient to realize the improvements that community participants sought in only about one-third of communities. In focus groups 6–12 months after the final program meeting, participants recalled that at least one of their efforts involving officials above the community level had realized their goals in 40% of communities in Ghana and 20% in Sierra Leone (Figure 4). But process tracing with interviews, observations, and focus groups suggests that these responses, combined with participants' locally focused efforts, contributed to participants in relatively more communities in Ghana and Sierra Leone eventually seeing their goals realized than in the program in Indonesia and Tanzania (Figure 5). Interviews with participants and others in their communities and observations of program meetings and of results of participants' efforts suggest that in 80% of communities in Ghana and Sierra Leone, at least one of participants' activities led to changes that achieved their objective. Observations of the final program meetings in Ghana suggest that these changes took longer: participants discussed at least one of their efforts as successful in the final program meeting in only 40% of communities in Ghana, compared with 100% of communities in Sierra Leone. Yet by the time of focus groups in both countries 6–12 months after the program ended, participants in most communities (80% in Ghana and 60% in Sierra Leone) recalled at least one of their activities achieving their objective, and in most communities (60% in both Ghana and Sierra Leone), participants were also able to recall specific, tangible changes that they saw as a result of their efforts, including facility transportation pools, road improvements, a delivery ward, new pumps to provide the facility with clean water, access to a functioning ambulance, and progress toward electrification (see Appendix D, Table D.4). In a similar proportion (60% of communities in Ghana and 80% in Sierra Leone), we were able to observe tangible changes from participants' activities. These proportions are substantially higher than in Indonesia and Tanzania. Analyses of responses to interviews with participants in Indonesia and Tanzania and others in their communities shortly after the program (Arkedis et al., forthcoming) suggest that participants' activities led to changes that clearly achieved their goal in less than half of communities (44% in Indonesia and 46% in Tanzania). In focus groups 1.5 years after the program, participants in approximately one-third (41% in Indonesia and 30% in Tanzania) recalled specific, tangible changes resulting from their efforts.

Figure 5. Efforts and outcomes

Proportion of communities where…

1) . . . participants in final program meeting discussed at least one activity being successful at meeting their objective*
2) . . . participants later recalled at least one planned activity as having been successful**
3) . . . participants later recalled their efforts as having generally improved health care**
4) . . . triangulating evidence suggests that at least one activity achieved participants' objective***
5) . . . participants later recalled tangible changes from their efforts**

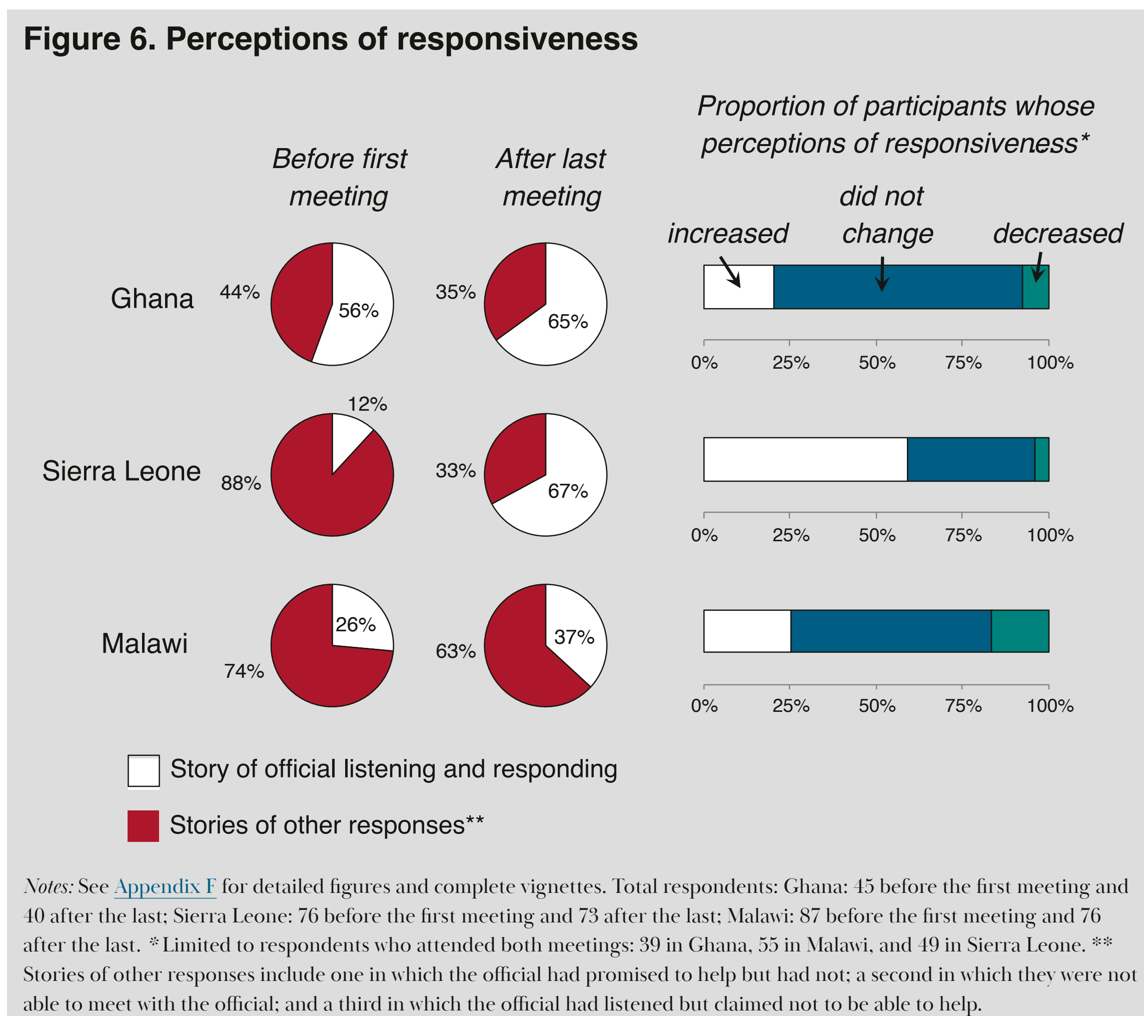


Sources: * meeting observations ** participant focus groups *** interviews and observations; does not include activities to raise community awareness and knowledge. See Appendix D and Arkedis et al. (forthcoming) for further discussion.

Interviews with officials after the program in Ghana and Sierra Leone also suggest that they generally found their interactions with community participants to be mutually constructive. Most (70% of officials interviewed) specifically recalled the meetings as opportunities to better understand what community members needed to improve their health care, to explain what they could and could not do to help, and to work with them around the shared goal of improving a valued public service. When asked to reflect on their participation in the program several officials recalled it as helping them with their responsibilities or making their jobs easier. As noted above, over the course of the program, several other officials became involved because they saw similar opportunities.

Finally, in both countries but particularly in Sierra Leone, interviews with community participants before and after the program suggest that collaborations with officials may have influenced average perceptions of government responsiveness among participants. In interviews prior to the start of the program and again after the final program meeting, we asked participants which of four anchoring vignettes about local officials was most similar to their experiences: one in which officials had listened and responded by trying to improve a public service and three others in which officials were remote or unable to help. Before the program began, 56% of participants in Ghana and 12% of participants in Sierra Leone said the story of an official listening and responding by trying to help was the most familiar. After the program, 21% of participants in Ghana described increased familiarity with responsive officials, similar proportionately to the increase among participants in Malawi (25%). In Sierra Leone, nearly six times as many participants chose the first story after the last meeting than did so before the first meeting. Figure 6 shows responses before and after the program; see Appendix F for the four vignettes.

Figure 6. Perceptions of responsiveness



Notes: See Appendix F for detailed figures and complete vignettes. Total respondents: Ghana: 45 before the first meeting and 40 after the last; Sierra Leone: 76 before the first meeting and 73 after the last; Malawi: 87 before the first meeting and 76 after the last. *Limited to respondents who attended both meetings: 39 in Ghana, 55 in Malawi, and 49 in Sierra Leone. ** Stories of other responses include one in which the official had promised to help but had not; a second in which they were not able to meet with the official; and a third in which the official had listened but claimed not to be able to help.

In short, the answer to our fourth question—whether more collaboration between community participants and officials contributed to more changes in maternal and newborn health care—is yes for the program in Ghana and Sierra Leone. The evidence described above does not imply that the second program was more likely than the first to have been sufficient to significantly raise health-care outcomes on average across all the communities where it was offered. In most communities where the second program was offered, participants saw only one tangible change, and both these changes and comparisons with the communities in the small control groups in each country suggest that none of the changes that they or officials recalled were encompassing reforms like policy changes that might have affected other communities (see Appendix E). But the comparisons above do suggest that participants in the program in Ghana and Sierra Leone saw tangible changes in their community's health care from their efforts in nearly twice as many communities as participants in the first program.

Overall, we find substantial differences between the experiences of participants in these two programs that are consistent with the observable implications in Section 4 of the hypothesis in Section 2. These include differences in how and how often community participants worked with officials, including on aspects of care with which officials could be helpful or over which they had formal authority, related to supply chains, facility staffing, and facility infrastructure; differences in how and how

often participants planned to involve officials in realizing their objectives; and differences in how often participants' efforts led to tangible changes that they and others in their communities noticed and remembered.

B. Malawi: A Community Scorecard Program Involving Traditional Authorities

In contrast, interviews and focus groups with participants and observations of program meetings in Malawi are inconsistent with the differences hypothesized in Section 2. All suggest that on the key observable implications in Table 3, participants' experiences in Malawi were far more similar to those of participants in Indonesia and Tanzania than to participants' in Ghana and Sierra Leone.

Community-focused efforts

Like participants in Indonesia and Tanzania, participants in Malawi focused primarily on steps they could take themselves or with others in their community, rather than with government officials (Table 6). Even in communities in Malawi where participants planned to try for supply-side improvements to health care, only 40% planned to involve officials, a proportion similar to the program in Indonesia and Tanzania (30% and 49% of communities, respectively) and substantially less than in the program in Ghana and Sierra Leone, where participants in all communities planned to involve officials in supply-side improvements.

The traditional authorities involved in the program in Malawi attended most of the meetings facilitators organized with community participants, and both Group Village Heads involved also met with participants outside these meetings. But traditional authorities in Malawi do not have the same degree of formal bureaucratic and administrative authority or access to resources as the officials involved in Ghana and Sierra Leone. One Group Village Head recalled speaking with several families about seeking prenatal care and offering to try to help community participants to talk with the Traditional Authority about arranging a meeting with a district health official if they wanted to improve the drug supply in their facility. Another Group Village Head recalled trying to help community participants revive their local Health Center Advisory Committee and encouraging community participants in their efforts to keep the health facility clean. In open-ended follow-up interviews one year after the program ended, only two of the five Traditional Authorities and Group Village Heads involved recalled more specific efforts to respond to participants' requests. Triangulating with other interviews and observations, we could weakly verify one: resolving a dispute between facility staff and the community.[46]

Few participants in Malawi recalled working with district officials with similar responsibilities and authority to those involved in the program in Ghana and Sierra Leone. In the one community where they did, their goal was limited: to clarify a concern about others in their community not seeking care in health facilities, community participants wanted to learn if the district government had any rules around care-seeking for community members who refused to go to the hospital because of religious beliefs. The official responded by informing them that although the government encourages people to seek medical care in facilities, everyone has the right not to go to the hospital if they choose.

Finally, unlike participants' reflections on their meetings with district officials in Ghana and Sierra Leone, several participants in the second interface meeting in Malawi, where participants were invited to discuss their plans and progress during a regularly scheduled meeting of the District Executive Committee, described the meeting as confusing and challenging. Officials and community participants were unsure of the participants' roles in the meeting, and participants' involvement was limited by language barriers as much of the District Executive Committee meeting was conducted in English. Officials in Ghana and Sierra Leone all had detailed recollections of the meetings and of how they had been involved in participants' efforts to improve their health care. In open-ended interviews in Malawi a year after the program, several traditional authorities and other officials who were involved in the District Executive Committee meeting had only vague recollections of meeting with community participants.

No tangible changes

Both our observations and participants' later reflections suggest that participants in Malawi did not see specific tangible changes from their efforts. In most communities, participants recalled their efforts contributing to pregnant women and new mothers seeking care more frequently at the facility, more husbands accompanying them, easier access to the facility on weekends, assistance at the facility even when they came without the recommended supplies for delivery, and several other changes that, as noted above, led them generally to think that their efforts overall had improved health care in their communities. But none recalled specific, tangible changes, such as a new ambulance or improvements to clinic infrastructure. This was unlike any of the other countries: in roughly one-third of communities in Indonesia and Tanzania and two-thirds of communities in Ghana and Sierra Leone participants later recalled seeing these kinds of tangible changes from their efforts.

Overall, interviews and focus groups also suggest more mixed experiences among participants in the program in Malawi than among participants in other countries. One year after the program, participants in all but one community in Malawi said that they were glad that they had participated.[47] Yet participants in Malawi were also the most likely across the five countries to remember personal costs from participating. In 80% of communities, a majority said that they had been treated warily in their communities or with the suspicion that they were being paid, or had experienced other personal costs. In one community, participants also mentioned facing a risk when they tried to pursue corruption: a threat, not carried out, of removal from the list of those in the community eligibility for subsidies.[48]

In short, we find support from several perspectives for the hypothesis in Section 2 about differences between the program in Ghana and Sierra Leone and the earlier program in Indonesia and Tanzania, and little support for this hypothesis in the program in Malawi. Qualitative interviews, observations, and focus groups all suggest that the scorecard program in Ghana and Sierra Leone led to more opportunities for mutually constructive collaboration between citizens and officials and to more improvements in more places than in the program in Indonesia and Tanzania. In Malawi, the additional meetings with traditional authorities brokered more mutually constructive state-society collaboration to improve public services. But like participants in the program in Indonesia and Tanzania, participants in Malawi focused mostly on what they could do as a community or with health providers to improve maternal and newborn health care, and saw fewer rather than more tangible changes to their community's health care than participants in Indonesia and Tanzania. Table 7 summarizes the consistencies and inconsistencies in these aspects of participants' experiences with the observable implications in Section 4 of the hypothesis in Section 2 about the difference between these two programs.

Table 7. Results expected and observed

	Community scorecard (large scale)		Community scorecard with officials (small scale)		
	Expected	Observed in Indonesia and Tanzania	Expected	Observed in Ghana and Sierra Leone	Observed in Malawi
Across diverse communities did community members participate meaningfully in the program and find the experience helpful for improving their community's health care?	Yes	Yes	Yes	Yes	Yes
Did participants in the program that involved officials work more extensively with officials?			More extensively than first program	More extensively than first program	Similar to first program
Did officials respond constructively?			Yes	Yes	Rarely**
Did more collaboration between community participants and officials contribute to more changes in maternal and newborn health care?			Yes	Yes, although not clear beyond the program period or for average health care*	No***
To what extent do these answers generalize across contexts?	Across diverse contexts, including health-care system development and political institutions. Potential scope conditions: long-improving health-care systems neither among world's most nor least developed.		Across diverse contexts, including health-care system development. Potential scope conditions: long-improving health-care systems and democratic political institutions.		

Notes: * The program in Ghana and Sierra Leone led to more collaboration between community participants and officials than in the program in Indonesia and Tanzania, improved perceptions of responsiveness, and led to more tangible changes related to health care; the evidence is not clear that collaboration endured beyond the period of the program or improved health care on average across all the communities where the program was offered. ** Participants in Malawi seldom tried to work with officials other than traditional authorities, although when they did the official responded. *** Collaboration between community participants and officials and tangible changes to health care were not more common in the program in Malawi than in the program in Indonesia and Tanzania.

6. Generalizability and Scope Conditions

The consistency in the findings about these two programs across varied contexts suggests several conclusions that may generalize. Interviews, observations, and focus groups suggest that in almost all communities across all five countries, community members took advantage of the opportunities the programs offered to discuss the state of their maternal and newborn health care, deliberate on how they could improve it, develop plans they could try, discuss these plans with others in their community, and reflect on their progress over several months as they tried these plans. Although some were disappointed and recalled wasted effort and personal cost, most participants recalled their experiences

as beneficial, and some were willing to continue their efforts months or years after the facilitator was no longer holding meetings. In four of the five countries, there were communities where participants' efforts were also associated with tangible changes memorable to them and other community members (approximately one-third of communities in Indonesia and Tanzania, and approximately two-thirds of communities in Ghana and Sierra Leone). In all five, the experience appeared to have sustained or improved most participants' perceptions of their capacities to improve their community.

The experiences of participants varied, but the differences discussed in Section 5 varied less with contextual differences between countries than with whether officials with formal authority who were willing to help were also involved. Section 5 noted a difference between the program in Ghana and in Sierra Leone that may have been partly associated with the relatively more developed health system and greater economic resources available in Ghana than in Sierra Leone (Figure 2): in Ghana officials often relied on government resources and in Sierra Leone more on efforts at resolving conflicts and connecting communities with nongovernmental support.[49] But otherwise, the interviews, focus groups, and observations described in the previous section suggest that differences in participants' experiences were inconsistent with the many contextual differences among these five countries. At the time, the health system in Ghana was far more similar in development and outcomes to those in Malawi and Tanzania than to the health system in Sierra Leone. Yet participants' experiences in Ghana and Sierra Leone were far more similar to each other than to participants' in the programs in Indonesia, Malawi, and Tanzania. In both Ghana and Sierra Leone, community participants worked more often with officials, more communities saw tangible changes, and they and the officials involved mostly recalled their collaborations as mutually constructive. At the time Indonesia's health system was more developed than Malawi's or Tanzania's. Yet participants in Indonesia shared experiences that were far more similar to participants' in Malawi and Tanzania than to participants' in Ghana and Sierra Leone.

These consistencies and differences in the experiences of participants in each program may not generalize to contexts without characteristics shared by these countries, including countries without democratic or rapidly democratizing political institutions, which may offer more opportunities for state-society collaboration, or without health systems that had already been the focus of decades of intensive resources and reform efforts, where there may be either substantially more or fewer opportunities for communities to find ways to further improve them. Indeed the evidence in Björkman and Svensson (2009) and Björkman Nyqvist, de Walque, and Svensson (2017) suggests that in the context of a less developed health system, dramatic and sustained improvement from a community scorecard program was more, not less, likely from efforts that did not involve officials or other third parties.

The similarities in the characteristics of the programs are also among the possible scope conditions on our findings, relative to the wider range of programmatic approaches to improving the transparency and accountability of public services studied in the literature. Our findings about the impacts of these programs on average health outcomes are consistent with critiques noted early in the paper about the sufficiency of programs that are local, that are focused on collaborative problem-solving, or that are not complemented with additional resources or efforts to improve state oversight and other capabilities, to bring about significant or transformational improvements. Both programs we consider here provided information and encouraged participation in community discussion, planning, and activities to improve access to quality maternal and newborn health-care services. Both programs were locally focused. Neither program offered additional resources or encouraged or incentivized any particular approach to improving health care. The findings discussed in this paper may not generalize to programs without these characteristics (Andrews, Pritchett, and Woolcock 2017; Ansell and Gash 2008; Baiocchi, Heller, and Silva 2011; Cornwall 2002, 2004; Fox 2007a, 2007b, 2015; Fung 2006a; Gaventa 2006; Grindle 2007, 1997; Lieberman, Posner, and Tsai 2014; McGee and Gaventa 2011; Olken 2007).

Overall, we conjecture that the differences we describe in the experiences of participants in Ghana and Sierra Leone from those of participants in Indonesia, Malawi, and Tanzania are less likely to reflect contextual differences than differences in the programs themselves: in particular, whether the programs 1) included additional meetings to broker collaboration with officials, and 2) involved officials with direct, formal responsibility for and authority over district public services. Several smaller differences discussed in Appendix B between the program in Malawi and in Ghana and Sierra Leone may also have played some role in the differences we observe, including in whether the interface meetings were scheduled separately from an existing forum (the District Executive Committee meeting), the length of time of the program, and the experience and familiarity of the nongovernmental organization with the district.[50] But many of the specific relative differences in the experiences of participants in Ghana and Sierra Leone from participants' experiences in Indonesia, Malawi, and Tanzania—including differences in whom participants worked with, how they tried to improve their community's health care, and whether they saw tangible changes—are directly related to whether officials were also involved and to whether these officials were elected and appointed officials formally responsible for district public services or traditional authorities whose influence was more informal and indirect.

7. Conclusion

This paper considers the hypothesis that a community scorecard program that includes meetings with officials and leaders can broker more mutually constructive collaboration between community members and officials to improve a public service. In particular, we ask whether a community scorecard program involving these meetings with officials led consistently, across diverse contexts, to more collaboration between community participants and officials that contributed to tangible changes in maternal and newborn health care in more communities than an otherwise similar community scorecard program that did not include similar meetings.

Evidence from interviews, observations, and focus groups offering comparable perspectives on participants' experiences in both programs consistently supports two conclusions about this hypothesis.

First, at least in these relatively democratic contexts, involving officials offered opportunities for mutually constructive collaboration between citizens and officials to improve public services. The evidence does not suggest that greater collaboration and change are a necessary result of the involvement of officials. In Malawi, where the leaders and officials involved had less responsibility and authority over district public services, participants' experiences were far more similar to those of participants in Indonesia and Tanzania, where facilitators did not organize meetings with officials. But in both Ghana and Sierra Leone, where the scorecard program included meetings between community participants and district-level officials with more responsibility and authority over district public services, participants worked with officials more often than in Indonesia and Tanzania and subsequently saw officials as more responsive, and both community participants and officials involved generally recalled their collaborations as mutually constructive. The evidence suggests that these collaborations complemented participants' community-based efforts and contributed to more communities seeing more tangible changes than in the program in Indonesia and Tanzania. These included a new water supply at a health facility, the opening of a new facility, and initial construction of another facility—changes that may have been far more challenging for community participants to realize without these collaborations with officials.

Second, differences in participants' experiences of the program in Ghana and Sierra Leone from those of participants in Malawi are more likely to reflect differences in the approach to involving officials in the program in Malawi than contextual differences. Contextual differences between Ghana and Sierra Leone are greater than between either and Malawi on several important dimensions, including

economic resources and the development of their health systems. Yet in both Ghana and Sierra Leone, where the program involved elected and appointed district officers with formal responsibility for and authority over district public services, participants worked with officials more often than in Malawi and substantially more communities saw tangible changes from their efforts. In both, the officials involved generally recalled participating as an opportunity to better understand what communities needed to improve their health care, to explain what they could and could not do to help, and to work with community members around the shared goal of improving a valued public service; in both, several recalled these collaborations as helping them with their responsibilities or making their jobs easier. As in Ghana and Sierra Leone, participants in Malawi, where the program primarily involved hereditary traditional authorities whose influence over district public services was more indirect and informal, often thought that their efforts had improved their communities' health care. But their experiences were far more similar to participants' in Indonesia and Tanzania than to participants' in Ghana and Sierra Leone.

These conclusions have important caveats. The evidence is qualitative. Despite their myriad other political, economic, cultural, social, and historical differences, the countries in which these programs were offered also shared important similarities, including health systems that had been improving for decades and, in all but one, relatively democratic political institutions. These similarities, as well as shared characteristics of the two programs, suggest two scope-condition hypotheses about the generalizability of these conclusions to other contexts. First, contexts where political institutions offer incentives for vertical accountability may be particularly likely to be contexts in which community scorecard programs that include officials and leaders in positions of influence and authority offer opportunities for brokering state-society collaboration. Second, average differences in the economic resources available to officials may influence the way officials try to work with community participants, including whether they rely on government resources or nongovernmental support.

Yet overall, the evidence suggests that at a time of growing international commitment to improving the transparency, accountability, responsiveness, and effectiveness of public services, involving officials with authority over public services in community scorecard programs can offer opportunities for brokering mutually constructive state-society collaboration around a shared goal of improving a valued public service—collaboration that both community participants and officials generally find to be mutually constructive, and that contributes in more communities to more tangible changes than programs without this brokered collaboration.

References

3ie. 2018. "Does Community-Driven Development Build Social Cohesion or Infrastructure?" Working paper brief (Governance). London: International Initiative for Impact Evaluation. https://www.3ieimpact.org/sites/default/files/2019-02/CDD-brief-wp30.pdf.

Ackerman, John. 2004. "Co-Governance for Accountability: Beyond 'Exit' and 'Voice.'" *World Development* 32 (3): 447–63. https://doi.org/10.1016/j.worlddev.2003.06.015.

Agrawal, Arun. 2005. "Environmentality: Community, Intimate Government, and the Making of Environmental Subjects in Kumaon, India." *Current Anthropology* 46 (2): 161–90. https://doi.org/10.1086/427122.

Aiyar, Yamini, Soumya Kapoor Mehta, and Salimah Samji. 2011. "Strengthening Public Accountability: Lessons from Implementing Social Audits in Andhra Pradesh." Working Paper. New Delhi: Accountability Initiative, Centre for Policy Research. http://stage-cprindia.gailabs.com/research/papers/strengthening-public-accountability-lessons-implementing-social-audits-andhra.

Alchian, Armen A., and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization." *American Economic Review* 62 (5): 777–95.

Andrews, Matt, Lant Pritchett, and Michael Woolcock. 2017. *Building State Capability: Evidence, Analysis, Action.* New York: Oxford University Press. https://bsc.cid.harvard.edu/building-state-capability-evidence-analysis-action.

Ansell, Chris, and Alison Gash. 2008. "Collaborative Governance in Theory and Practice." *Journal of Public Administration Research and Theory* 18 (4): 543–71. https://doi.org/10.1093/jopart/mum032.

Arkedis, Jean, Jessica Creighton, Akshay Dixit, Archon Fung, Stephen Kosack, Dan Levy, and Courtney Tolmie. Forthcoming. "Can Transparency and Accountability Programs Improve Health? Experimental Evidence from Indonesia and Tanzania." *World Development.*

Armah-Attoh, Daniel, Edem Selormey, and Richard Houessou. 2016. "Despite Gains, Barriers Keep Health Care High on Africa's Priority List." Afrobarometer Policy Paper No. 31. https://www.afrobarometer.org/publications/pp31-despite-gains-barriers-keep-health-care-high-on-africas-priority-list.

Baiocchi, Gianpaolo, Patrick Heller, and Marcelo K. Silva. 2011. *Bootstrapping Democracy: Transforming Local Governance and Civil Society in Brazil.* Palo Alto: Stanford University Press.

Bandura, Albert. 1977. "Self-Efficacy: Toward a Unifying Theory of Behavioral Change." *Psychological Review* 84 (2): 191–215. https://doi.org/10.1037/0033-295X.84.2.191.

———. 1982. "Self-Efficacy Mechanism in Human Agency." *American Psychologist* 37 (2): 122–47. https://doi.org/10.1037/0003-066X.37.2.122.

Banerjee, Abhijit V., Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani. 2006. "Can Information Campaigns Spark Local Participation and Improve Outcomes? A Study of Primary Education in Uttar Pradesh, India." World Bank Policy Research Working Paper No. 3967. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=923274.

Banerjee, Abhijit V., Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani. 2010. "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India." *American Economic Journal: Economic Policy* 2 (1): 1–30. https://doi.org/10.1257/pol.2.1.1.

Banerjee, Abhijit V., Rema Hanna, Jordan Kyle, Benjamin A. Olken, and Sudarno Sumarto. 2018. "Tangible Information and Citizen Empowerment: Identification Cards and Food Subsidy Programs in Indonesia." *Journal of Political Economy* 126 (2): 451–91. https://doi.org/10.1086/696226.

Basurto, Maria Pia, Pascaline Dupas, and Jonathan Robinson. 2020. "Decentralization and Efficiency of Subsidy Targeting: Evidence from Chiefs in Rural Malawi." *Journal of Public Economics* 185: 104047. https://doi.org/10.1016/j.jpubeco.2019.07.006.

Besley, Timothy, and Torsten Persson. 2010. "State Capacity, Conflict, and Development." *Econometrica* 78 (1): 1–34. https://doi.org/10.3982/ECTA8073.

Björkman Nyqvist, Martina, Darin Christensen, Archon Fung, Stephen Kosack, and Pia Raffler. 2019. "Transparency and Accountability Programs: Transformative, or Irrelevant?" Roundtable session at the 115th American Political Science Association's Annual Meeting & Exhibition, Washington, DC.

Björkman, Martina, and Jakob Svensson. 2009. "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda." *The Quarterly Journal of Economics* 124 (2): 735–69. https://doi.org/10.1162/qjec.2009.124.2.735.

Björkman Nyqvist, Martina, Damien de Walque, and Jakob Svensson. 2017. "Experimental Evidence on the Long-Run Impact of Community-Based Monitoring." *American Economic Journal: Applied Economics* 9 (1): 33–69. https://doi.org/10.1257/app.20150027.

Brady, Henry E. 2008. "Causation and Explanation in Social Science." In *The Oxford Handbook of Political Methodology*, edited by Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier. New York: Oxford University Press. http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199286546.001.0001/oxfordhb-9780199286546-e-10.

Bratton, Michael, and E. Gyimah-Boadi. 2016. "Do Trustworthy Institutions Matter for Development? Corruption, Trust, and Government Performance in Africa." Afrobarometer Dispatch No. 112. Afrobarometer. https://www.afrobarometer.org/publications/ad112-do-trustworthy-institutions-matter-development-corruption-trust-and-government.

Brixi, Hana, Ellen Lust, and Michael Woolcock. 2015. *Trust, Voice, and Incentives: Learning from Local Success Stories in Service Delivery in the Middle East and North Africa.* Washington, DC: The World Bank. https://openknowledge.worldbank.org/handle/10986/21607.

Carlson, Elizabeth, and Brigitte Seim. 2020. "Honor among Chiefs: An Experiment on Monitoring and Diversion Among Traditional Leaders in Malawi." *Journal of Development Studies* 56 (8): 1541–57. https://doi.org/10.1080/00220388.2019.1703955.

Chambers, Robert. 1997. *Whose Reality Counts? Putting the First Last.* London: Intermediate Technology Publications.

Christensen, Darin, Oeindrila Dube, Johannes Haushofer, Bilal Siddiqi, and Maarten Voors. 2020. "Building Resilient Health Systems: Experimental Evidence from Sierra Leone and the 2014 Ebola Outbreak." *Quarterly Journal of Economics.* https://doi.org/10.1093/qje/qjaa039.

Collier, David. 1993. "The Comparative Method." In *Political Science: The State of the Discipline II*, edited by Ada W. Finifter, 105–19. Washington, DC: American Political Science Association.

Cornwall, Andrea. 2002. "Making Spaces, Changing Places: Situating Participation in Development." IDS Working Paper 170. Sussex: Institute of Development Studies. https://www.ids.ac.uk/publications/making-spaces-changing-places-situating-participation-in-development/.

———. 2004. "New Democratic Spaces? The Politics and Dynamics of Institutionalised Participation." *IDS Bulletin* 35 (2). https://doi.org/10.1111/j.1759-5436.2004.tb00115.x.

Creighton, Jessica, Jean Arkedis, Archon Fung, Stephen Kosack, Dan Levy, and Courtney Tolmie. 2020. "Insights from Transparency and Accountability Action Plans in Indonesia and Tanzania." Cambridge, MA: Ash Center for Democratic Governance and Innovation, Harvard Kennedy School. https://ash.harvard.edu/publications/insights-transparency-and-accountability-action-plans-indonesia-and-tanzania.

Donahue, John D., and Richard J. Zeckhauser. 2012. *Collaborative Governance: Private Roles for Public Goals in Turbulent Times.* Princeton, NJ: Princeton University Press.

Evans, Peter. 1995. *Embedded Autonomy: States and Industrial Transformation.* Princeton, NJ: Princeton University Press.

———. 1996. "Government Action, Social Capital and Development: Reviewing the Evidence on Synergy." *World Development* 24 (6): 1119–32. https://doi.org/10.1016/0305-750X(96)00021-6.

Farmer, Paul, Jim Yong Kim, Arthur Kleinman, and Matthew Basilico, eds. 2013. *Reimagining Global Health: An Introduction.* Berkeley: University of California Press. https://www.ucpress.edu/book/9780520271999/reimagining-global-health.

Fishkin, James S., and Robert C. Luskin. 2005. "Experimenting with a Democratic Ideal: Deliberative Polling and Public Opinion." *Acta Politica* 40: 284–98. https://doi.org/10.1057/palgrave.ap.5500121.

Fox, Jonathan. 2007a. *Accountability Politics: Power and Voice in Rural Mexico.* New York: Oxford University Press.

———. 2007b. "The Uncertain Relationship between Transparency and Accountability." *Development in Practice* 17 (4–5): 663–71. https://doi.org/10.1080/09614520701469955.

———. 2015. "Social Accountability: What Does the Evidence Really Say?" *World Development* 72: 346–361. https://doi.org/10.1016/j.worlddev.2015.03.011.

Freedom House. 2020. "Freedom in the World — Tanzania Country Report." Washington, DC: Freedom House.

Fung, Archon. 2006a. *Empowered Participation: Reinventing Urban Democracy*. Princeton, NJ: Princeton University Press. https://press.princeton.edu/titles/7762.html.

——. 2006b. "Varieties of Participation in Complex Governance." *Public Administration Review* 66 (1): 66–75. https://doi.org/10.1111/j.1540-6210.2006.00667.x.

Fung, Archon, and Erik Olin Wright. 2001. "Deepening Democracy: Innovations in Empowered Participatory Governance." *Politics and Society* 29 (1): 5–41. https://doi.org/10.1177/0032329201029001002.

——. 2003. *Deepening Democracy: Institutional Innovations in Empowered Participatory Governance*. London: Verso.

Gaventa, John. 2006. "Finding the Spaces for Change: A Power Analysis." *IDS Bulletin* 37 (6): 23–33. https://doi.org/10.1111/j.1759-5436.2006.tb00320.x.

Giraudy, Agustina, Eduardo Moncada, and Richard Snyder, eds. 2019. *Inside Countries: Subnational Research in Comparative Politics*. Cambridge: Cambridge University Press.

Grindle, Merilee S. 1997. *Getting Good Government: Capacity Building in the Public Sectors of Developing Countries*. Cambridge, MA: Harvard Institute for International Development.

——. 2007. *Going Local: Decentralization, Democratization, and the Promise of Good Governance*. Princeton, NJ: Princeton University Press.

Grossman, Gene M., and Elhanan Helpman. 2001. *Special Interest Politics*. Cambridge, MA: MIT Press.

Gullo, Sara, Christine Galavotti, and Lara Altman. 2016. "A Review of CARE's Community Score Card Experience and Evidence." *Health Policy and Planning* 31 (10): 1467–78. https://doi.org/10.1093/heapol/czw064.

Hernández, Alison, Ana Lorena Ruano, Anna-Karin Hurtig, Isabel Goicolea, Miguel San Sebastián, and Walter Flores. 2019. "Pathways to Accountability in Rural Guatemala: A Qualitative Comparative Analysis of Citizen-Led Initiatives for the Right to Health of Indigenous Populations." *World Development* 113: 392–401. https://doi.org/10.1016/j.worlddev.2018.09.020.

Hirschman, Albert O. 1970. *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Cambridge, MA: Harvard University Press.

Holland, Jeremy, and Florian Schatz. 2016. *Macro Evaluation of DFID's Policy Frame for Empowerment and Accountability*. London: ITAD. https://itad.com/wp-content/uploads/2017/06/EA-Macro-Evaluation-Technical-report-Dec16-FINAL.pdf.

Hsia, Renee Y., Naboth A. Mbembati, Sarah Macfarlane, and Margaret E. Kruk. 2012. "Access to Emergency and Surgical Care in Sub-Saharan Africa: The Infrastructure Gap." *Health Policy and Planning* 27 (3): 234–44. https://doi.org/10.1093/heapol/czr023.

Jenkins, Rob, and Anne Marie Goetz. 1999. "Accounts and Accountability: Theoretical Implications of the Right-to-Information Movement in India." *Third World Quarterly* 20 (3): 603–22. https://doi.org/10.1080/01436599913712.

Joshi, Anuradha. 2008. "Producing Social Accountability? The Impact of Service Delivery Reforms." *IDS Bulletin* 38 (6): 10–17. https://doi.org/10.1111/j.1759-5436.2007.tb00414.x.

——. 2010. "Review of Impact and Effectiveness of Transparency and Accountability Initiatives. Annex 1: Service Delivery." Institute of Development Studies. https://www.gov.uk/research-for-development-outputs/review-of-impact-and-effectiveness-of-transparency-and-accountability-initiatives-annex-1-service-delivery.

Joshi, Anuradha, and Peter P. Houtzager. 2012. "Widgets or Watchdogs? Conceptual Explorations in Social Accountability." *Public Management Review* 14 (2): 145–62. https://doi.org/10.1080/14719037.2012.657837.

J-PAL. 2011. "Governance Review Paper: JPAL Governance Initiative." Abdul Latif Jameel Poverty Action Lab, MIT. https://www.povertyactionlab.org/sites/default/files/documents/J-PAL_GI-ReviewPaper_vMay-2011.pdf.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47 (2): 263–91. https://doi.org/10.2307/1914185.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.

King, Gary, Christopher J. L. Murray, Joshua A. Salomon, and Ajay Tandon. 2004. "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research." *American Political Science Review* 98 (1): 191–207. https://doi.org/10.1017/S000305540400108X.

Kosack, Stephen, Grant Bridgman, Jessica Creighton, Courtney Tolmie, and Archon Fung. 2019. "Encouraged Participation." Working Paper. https://ash.harvard.edu/files/ash/files/encouraging_participation.pdf.

Kosack, Stephen, and Archon Fung. 2014. "Does Transparency Improve Governance?" *Annual Review of Political Science* 17: 65–87. https://doi.org/10.1146/annurev-polisci-032210-144356.

Kruk, Margaret E., Anna D. Gage, Catherine Arsenault, Keely Jordan, Hannah H. Leslie, Sanam Roder-DeWan, Olusoji Adeyi, Pierre Barker, Bernadette Daelmans, Svetlana V. Doubova, Mike English, Ezequiel García Elorrio, Frederico Guanais, Oye Gureje, Lisa R. Hirschhorn, Lixin Jiang, Edward Kelley, Ephrem Tekle Lemango, Jerker Liljestrand, Address Malata, Tanya Marchant, Malebona Precious Matsoso, John G. Meara, Manoj Mohanan, Youssoupha Ndiaye, Ole F. Norheim, K. Srinath Reddy, Alexander K. Rowe, Joshua A. Salomon, Gagan Thapa, Nana A.Y. Twum-Danso, and Muhammad Pate. 2018. "High-Quality Health Systems in the Sustainable Development Goals Era: Time for a Revolution." *The Lancet Global Health* 6 (11): E1196–E1252. https://doi.org/10.1016/S2214-109X(18)30386-3.

Kruk, Margaret E., Hannah H. Leslie, Stéphane Verguet, Godfrey M. Mbaruku, Richard M. K. Adanu, and Ana Langer. 2016. "Quality of Basic Maternal Care Functions in Health Facilities of Five African Countries: An Analysis of National Health System Surveys." *The Lancet Global Health* 4 (11): E845–55. https://doi.org/10.1016/S2214-109X(16)30180-2.

Levitsky, Steven, and Lucan A. Way. 2010. *Competitive Authoritarianism: Hybrid Regimes After the Cold War.* New York: Cambridge University Press.

Lieberman, Evan S. 2005. "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review* 99 (3): 435–52. https://doi.org/10.1017/S0003055405051762.

Lieberman, Evan S., Daniel N. Posner, and Lily L. Tsai. 2014. "Does Information Lead to More Active Citizenship? Evidence from an Education Intervention in Rural Kenya." *World Development* 60: 69–83. https://doi.org/10.1016/j.worlddev.2014.03.014.

Lijphart, Arend. 1971. "Comparative Politics and the Comparative Method." *American Political Science Review* 65 (3): 682–93.

Lipsky, Michael. 1980. *Street Level Bureaucracy: Dilemmas of the Individual in Public Services.* New York: Russell Sage.

Mahoney, James. 2008. "Toward a Unified Theory of Causality." *Comparative Political Studies* 41 (4/5): 412–36. https://doi.org/10.1177/0010414007313115.

Mainwaring, Scott, and Christopher Welna, eds. 2003. *Democratic Accountability in Latin America.* Oxford: Oxford University Press.

Mansuri, Ghazala, and Vijayendra Rao. 2013. *Localizing Development: Does Participation Work?* Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/11859.

Marathe, Shweta, Deepali Takkundi, Trupti Malti, and Abhay Shukla. 2020. "Research Brief on the Impact of the Community Based Monitoring and Planning Process in the State of Maharashtra." Support for Advocacy and Training to Health Initiatives. http://sathicehat.org/wp-content/uploads/2020/02/Research-Brief-on-the-impact-of-the-Community-Based-Monitoring-and-Planning-process-in-the-state-of-Maharashtra.pdf.

Marshall, Monty G., and Ted Robert Gurr. 2014. "Polity IV Project." Center for Systemic Peace. https://www.systemicpeace.org/polity/polity4.htm.

Masset, Edoardo. 2015. "Measuring Empowerment in Rural India Using Vignettes." *Journal of Development Effectiveness* 7 (3): 346–56. https://doi.org/10.1080/19439342.2015.1068355.

McGee, Rosie, and John Gaventa. 2011. "Shifting Power? Assessing the Impact of Transparency and Accountability Initiatives." IDS Working Paper 383: 1–39. https://doi.org/10.1111/j.2040-0209.2011.00383_2.x.

Migdal, Joel S. 2001. *State in Society: Studying How States and Societies Transform and Constitute One Another.* New York: Cambridge University Press.

Mill, John Stuart. 1843. *A System of Logic.* London.

Moe, Terry M. 1988. *The Organization of Interests: Incentives and the Internal Dynamics of Political Interest Groups.* Chicago: University of Chicago Press.

North, Douglass Cecil. 1990. *Institutions, Institutional Change, and Economic Performance.* New York: Cambridge University Press.

OECD. 2008. *Paris Declaration on Aid Effectiveness (2005); Accra Agenda for Action (2008).* Paris: OECD. http://www.oecd.org/dac/effectiveness/34428351.pdf.

Olken, Benjamin A. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115 (2): 200–249. https://doi.org/10.1086/517935.

Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups.* Cambridge, MA: Harvard University Press.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action.* New York: Cambridge University Press.

———. 1996. "Crossing the Great Divide: Coproduction, Synergy, and Development." *World Development* 24 (6): 1073–87. https://doi.org/10.1016/0305-750X(96)00023-X.

Pew Research Center. 2017. "Globally, Broad Support for Representative and Direct Democracy." https://www.pewresearch.org/global/2017/10/16/globally-broad-support-for-representative-and-direct-democracy/.

Przeworski, Adam, Michael E. Alvarez, José Antonio Cheibub, and Fernando Limongi. 2000. *Democracy and Development: Political Institutions and Well-Being in the World, 1950–1990.* Cambridge: Cambridge University Press.

Putnam, Robert D., Robert Leonardi, and Raffaella Y. Nanetti. 1994. *Making Democracy Work: Civic Traditions in Modern Italy.* Princeton, NJ: Princeton University Press.

Raffler, Pia, Daniel N. Posner, and Doug Parkerson. 2019. "The Weakness of Bottom-Up Accountability: Experimental Evidence from the Ugandan Health Sector." Working Paper. https://www.poverty-action.org/sites/default/files/publications/The%20Weakness%20of%20Bottom-Up%20Accountability.pdf.

Sabel, Charles F. 1993. "Studied Trust: Building New Forms of Cooperation in a Volatile Economy." *Human Relations* 46 (9): 1133–70. https://doi.org/10.1177/001872679304600907.

Sachs, Jeffrey D. 2005. *The End of Poverty: Economic Possibilities for Our Time.* New York: Penguin.

Seawright, Jason. 2016. *Multi-Method Social Science: Combining Qualitative and Quantitative Tools.* Cambridge: Cambridge University Press.

Sen, Amartya. 1999. *Development as Freedom.* New York: Alfred A. Knopf.

Snyder, Richard. 2001. "Scaling Down: The Subnational Comparative Method." *Studies in Comparative International Development* 36 (1): 93–110. https://doi.org/10.1007/BF02687586.

Souza, Celina. 2001. "Participatory Budgeting in Brazilian Cities: Limits and Possibilities in Building Democratic Institutions." *Environment and Urbanization* 13 (1): 159–84. https://doi.org/10.1177/095624780101300112.

Stern, Elliot, Nicoletta Stame, John Mayne, Kim Forss, Rick Davies, and Barbara Befani. 2012. "Broadening the Range of Designs and Methods for Impact Evaluations." Working Paper 38. London: DFID. https://www.oecd.org/derec/50399683.pdf.

Thampi, Gopa Kumar. 2017. "Social Accountability from the Trenches: 6 Critical Reflections." *From Poverty to Power* (blog), May 3. https://oxfamblogs.org/fp2p/social-accountability-from-the-trenches-6-critical-reflections/.

Transparency for Development. 2015. "Evaluation Design Report." Cambridge, MA: Ash Center for Democratic Governance and Innovation, Harvard Kennedy School. https://ash.harvard.edu/files/ash/files/evaluation_design_report.pdf.

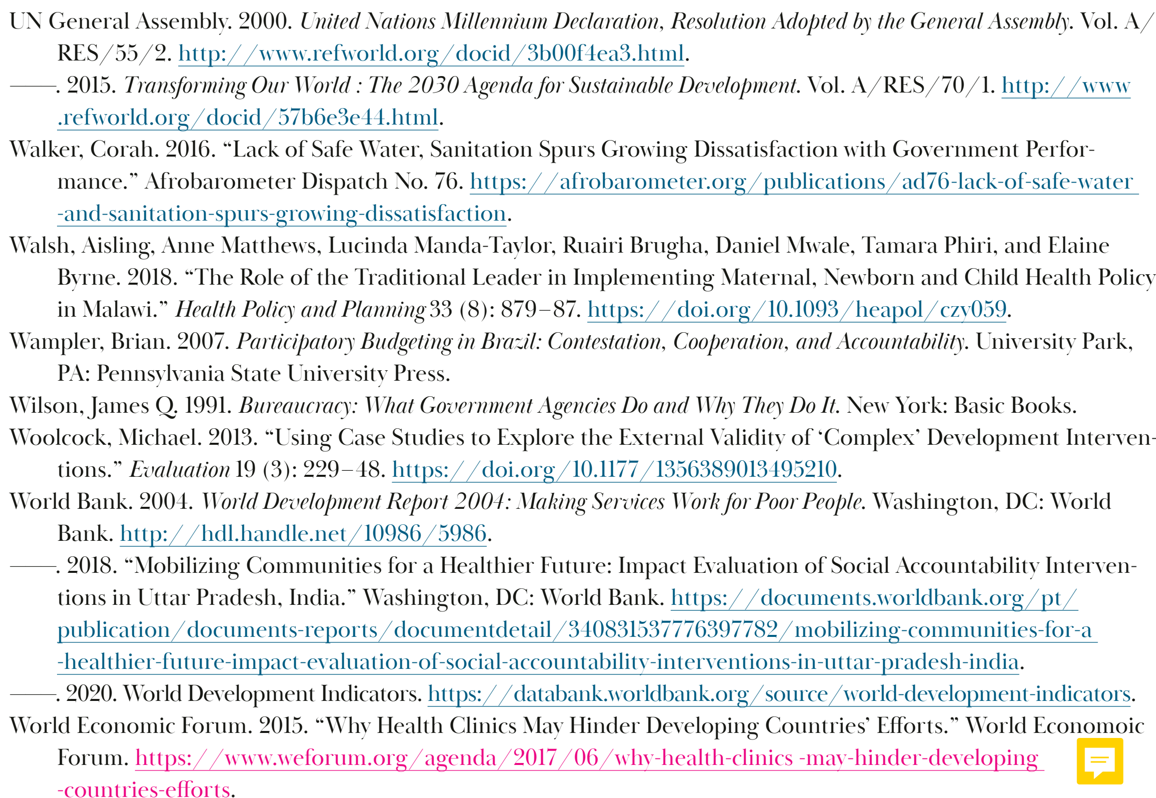
———. 2016. "Baseline Report." Cambridge, MA: Ash Center for Democratic Governance and Innovation, Harvard Kennedy School. https://ash.harvard.edu/files/ash/files/baseline_report.pdf.

———. 2017. "Citizen Voices, Community Solutions: Designing Better Transparency and Accountability Approaches to Improve Health." Ash Center Occasional Papers. Cambridge, MA: Ash Center for Democratic Governance and Innovation, Harvard Kennedy School. https://ash.harvard.edu/files/ash/files/citizen_voices_community_solutions.pdf.

Tsebelis, George. 2002. *Veto Players: How Political Institutions Work.* Princeton, NJ: Princeton University Press.

Tversky, Amos, and Daniel Kahneman. 1991. "Loss Aversion in Riskelss Choice: A Reference-Dependent Model." *Quarterly Journal of Economics* 106 (4): 1039–61. https://doi.org/10.2307/2937956.

UNDP. 1990. *Human Development Report 1990: Concept and Measurement of Human Development.* New York: Oxford University Press for the United Nations Development Program. http://www.hdr.undp.org/en/reports/global/hdr1990.



UN General Assembly. 2000. *United Nations Millennium Declaration, Resolution Adopted by the General Assembly*. Vol. A/RES/55/2. http://www.refworld.org/docid/3b00f4ea3.html.

——. 2015. *Transforming Our World : The 2030 Agenda for Sustainable Development*. Vol. A/RES/70/1. http://www.refworld.org/docid/57b6e3e44.html.

Walker, Corah. 2016. "Lack of Safe Water, Sanitation Spurs Growing Dissatisfaction with Government Performance." Afrobarometer Dispatch No. 76. https://afrobarometer.org/publications/ad76-lack-of-safe-water-and-sanitation-spurs-growing-dissatisfaction.

Walsh, Aisling, Anne Matthews, Lucinda Manda-Taylor, Ruairi Brugha, Daniel Mwale, Tamara Phiri, and Elaine Byrne. 2018. "The Role of the Traditional Leader in Implementing Maternal, Newborn and Child Health Policy in Malawi." *Health Policy and Planning* 33 (8): 879–87. https://doi.org/10.1093/heapol/czy059.

Wampler, Brian. 2007. *Participatory Budgeting in Brazil: Contestation, Cooperation, and Accountability*. University Park, PA: Pennsylvania State University Press.

Wilson, James Q. 1991. *Bureaucracy: What Government Agencies Do and Why They Do It*. New York: Basic Books.

Woolcock, Michael. 2013. "Using Case Studies to Explore the External Validity of 'Complex' Development Interventions." *Evaluation* 19 (3): 229–48. https://doi.org/10.1177/1356389013495210.

World Bank. 2004. *World Development Report 2004: Making Services Work for Poor People*. Washington, DC: World Bank. http://hdl.handle.net/10986/5986.

——. 2018. "Mobilizing Communities for a Healthier Future: Impact Evaluation of Social Accountability Interventions in Uttar Pradesh, India." Washington, DC: World Bank. https://documents.worldbank.org/pt/publication/documents-reports/documentdetail/340831537776397782/mobilizing-communities-for-a-healthier-future-impact-evaluation-of-social-accountability-interventions-in-uttar-pradesh-india.

——. 2020. World Development Indicators. https://databank.worldbank.org/source/world-development-indicators.

World Economic Forum. 2015. "Why Health Clinics May Hinder Developing Countries' Efforts." World Economoic Forum. https://www.weforum.org/agenda/2017/06/why-health-clinics -may-hinder-developing-countries-efforts.

Introduction to the Appendixes

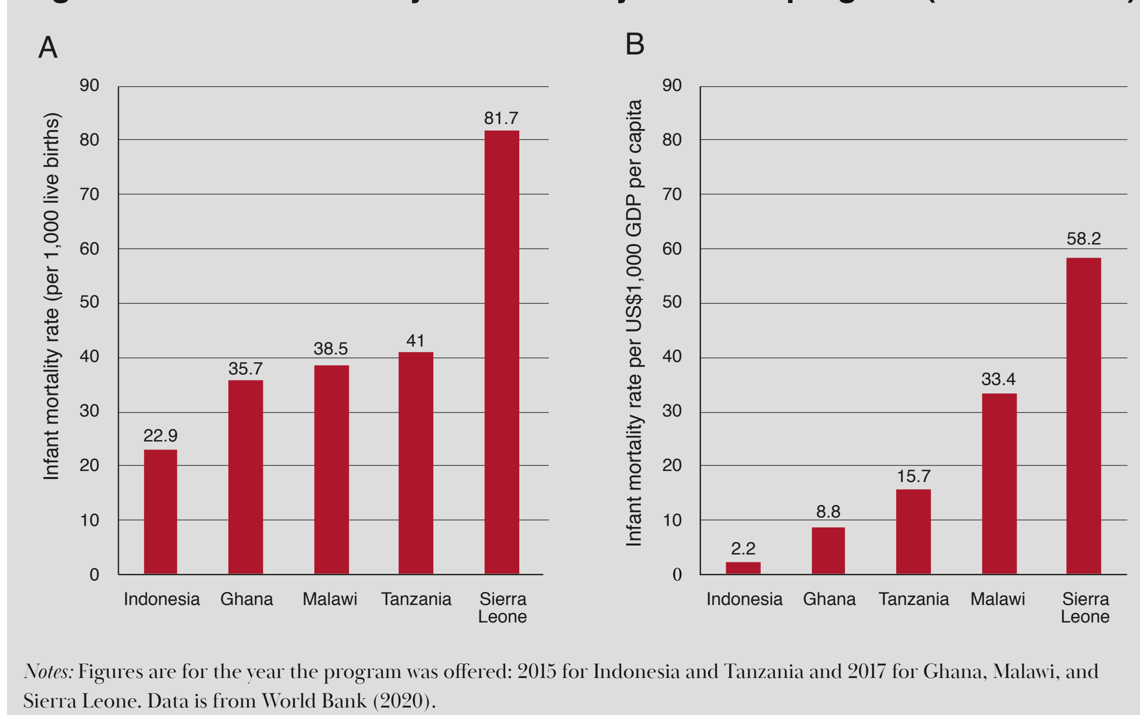
Appendixes A and B supplement the brief discussion in Sections 3 and 4 of the paper with further detail: Appendix A on contexts where the two programs were offered, including infant mortality rates and political institutions in the period immediately before them, and Appendix B on the two programs themselves. Appendix C details activities participants planned, whom they tried to involve in their efforts, and their progress in attempting and completing what they had planned. Appendix D expands on the discussion in the paper of participants' perceptions of the efficacy of their efforts. Appendix E includes more detailed descriptions of officials' responses to participants' efforts. Appendix F includes scripts for the anchoring vignettes used in interviews with participants before and after the program.

Appendix A. Contexts

Ghana, Indonesia, Malawi, Sierra Leone, and Tanzania differ economically, politically, culturally, geographically, and in many other ways, offering a widely diverse group of country contexts in which to identify consistent patterns relevant to the hypothesis in Section 2 of the main paper. But all five also share long-improving health systems, and four of the five share relatively democratic political institutions—two relative similarities of potential relevance to the generalizability and scope conditions of our findings about the hypothesis (see Section 3 of the main paper). Here we offer additional discussion of variation in the development of their health systems as well as additional discussion of the two similarities.

Aside from Sierra Leone, which has one of the world's highest infant mortality rates, none of the health systems in these five countries were among the world's most or least developed: all offered lifesaving care, yet none were universally accessible, and in all, pregnant women and infants regularly suffered and died from preventable causes. Yet at the time the programs were offered, health outcomes varied widely among the five countries. In 2015, infant mortality rates in Tanzania (41 per 1,000 births) were nearly double those in Indonesia (23 per 1,000 births). In 2017, infant mortality rates in Sierra Leone (82 per 1,000 births) were more than double those in Ghana and Malawi (36 in Ghana, 39 in Malawi). Figure A.1, Panel A, supplements Figure 1 in the main paper with infant (under one year) mortality rates in the initial year of the program, 2015 for the program in Indonesia and Tanzania or 2017 for the program in Ghana, Malawi, and Sierra Leone. Panel B shows infant mortality rates per $1,000 of gross domestic product (GDP) per capita, to adjust for the varying levels of economic resources available in each country.

Figure A.1. Infant mortality rates in the year of the program (2015 or 2017)



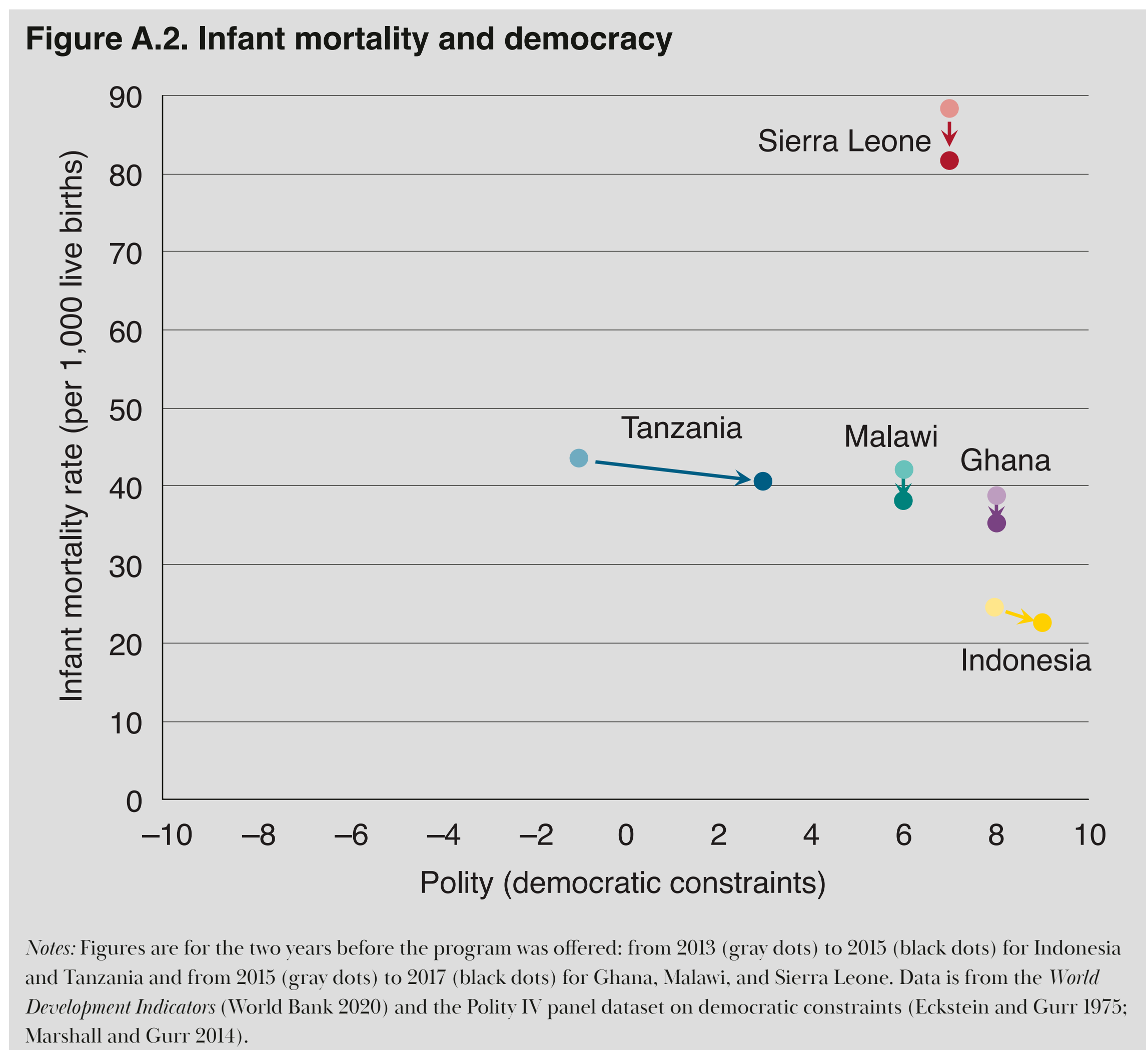
Notes: Figures are for the year the program was offered: 2015 for Indonesia and Tanzania and 2017 for Ghana, Malawi, and Sierra Leone. Data is from World Bank (2020).

The five countries also share similarities on two important contextual dimensions (see Section 3 of the paper). First, health-care systems in all five have long been the focus of intensive resources and reforms, have been improving steadily for decades, and continued to improve during and after the two scorecard programs.[51] In all five, improving health was also a particular focus for the current government, and all were in the midst of major health-care reforms and new investment. At the time, Indonesia was overhauling its national health-insurance scheme to increase access to health care (Mboi 2015; Maharani et al. 2019; Erlangga, Ali, and Bloor 2019). In the 2017–2018 fiscal year the Tanzanian government increased its health-sector budget by 28%, to 7% of the national budget and 1.8% of GDP (UNICEF 2018). In the years before the scorecard program, the national governments had been expanding the health systems of Ghana (Escribano-Ferrer et al. 2016; Ghana Health Service 2018) and Malawi (Chansa and Pattnaik 2018). Health indicators including maternal and newborn health were also improving in Sierra Leone, following efforts to restore and strengthen the system after the devastation from the Ebola outbreak in 2014 (WHO 2017). As we note in the main paper, this past and ongoing progress may have also steadily reduced remaining unrealized opportunities for communities using these services to further improve them without support from officials, NGOs, and other third parties, one reason a community scorecard program might have had more impact on outcomes in a system like Uganda's in 2004 than Uganda's in 2015 (Björkman and Svensson 2009; Björkman Nyqvist, de Walque, and Svensson 2017; Raffler, Posner, and Parkerson 2019).

Second, four of the five countries shared a similarity that existing theory suggests could be important to the opportunities a community scorecard program might offer for collaborative problem-solving with officials: national institutional incentives for vertical democratic accountability.[52] The second program involving officials was purposefully offered in places with this similarity. The institutions shaping these incentives were recent in all three. Ghana, Malawi, and Sierra Leone are all among the world's younger democracies. All are postcolonial states where prior state-society relations had not always been collaborative and trusting and where government was seen by many as relatively remote from citizens' needs. Yet all three, as well as Indonesia, also shared more recent experience with regularly contested democratic elections and multiple peaceful transfers of national political power. Levitsky and Way's (2010) commonly used definitions distinguish democratic nation-states from competitive authoritarian and fully authoritarian states as those with more than 5 on the −10 to +10 Polity index, a measure of democratic constraints on executive authority (Eckstein and Gurr 1975, vii, 488; Marshall and Gurr 2014). On this metric, in the year of the program Ghana (+8), Indonesia (+9), Malawi (+6), and Sierra Leone (+7) were all democracies. All had been through transfers of national political power in the years before the program: 2016 in Ghana, when power shifted from John Mahama of the National Democratic Congress to Nana Akufo-Addo of the opposition New Patriotic Party, the country's third consecutive transfer of presidential power between opposing political parties; 2014 in Indonesia with the election of Joko Widodo, who defeated a candidate supported by the party of the previous incumbent who was barred from seeking another term, Susilo Bambang Yudhoyono; 2018 in Sierra Leone, where Julius Maada Bio of the Sierra Leone People's Party (SLPP), previously leader of the opposition, won the presidency while the All People's Congress (APC), which had previously held the presidency, maintained a majority in the parliament; and 2014 in Malawi, where Peter Mutharika, the candidate of the opposition Democratic Progressive Party (DPP), defeated the incumbent of the People's Party, Joyce Banda, in a four-way race with two other parties—the Malawi Congress Party (MCP) and the United Democratic Front. The exception among the five countries is Tanzania: although at the time of the program Tanzania had a strong opposition party, Chadema, and political institutions that were democratizing, it was still a "competitive authoritarian" state, in which one party, Chama Cha Mapinduzi (CCM), currently led by John Magufuli, has retained power for more than half a century.[53]

Figure A.2 plots infant mortality rates for these five countries in the two years prior to the program together with the Polity IV measure of democratic constraints on executive authority for each.

Figure A.2. Infant mortality and democracy



Notes: Figures are for the two years before the program was offered: from 2013 (gray dots) to 2015 (black dots) for Indonesia and Tanzania and from 2015 (gray dots) to 2017 (black dots) for Ghana, Malawi, and Sierra Leone. Data is from the *World Development Indicators* (World Bank 2020) and the Polity IV panel dataset on democratic constraints (Eckstein and Gurr 1975; Marshall and Gurr 2014).

In addition to these two contextual similarities, all five countries where these programs were offered were purposefully selected to be similar in a third respect: each had a local civil-society organization with substantial local knowledge and experience both of the local context and of working on similar programs, along with a willingness to actively engage in the co-design process described in Section 3 of the paper and implement the program.

Communities

The 215 communities where these two programs were offered were randomly selected from among communities in the catchment areas of health facilities stratified across two regions of Tanzania, two provinces of Indonesia, and one district each in Ghana, Malawi, and Sierra Leone. A second group of 209 other communities were randomly selected from among the catchment areas of other health facilities to be part of a control group, similarly stratified across the same regions, provinces, and districts so that at baseline the two groups of communities were similar in their access to maternal and newborn health care.

In Indonesia, communities in the two groups were randomly selected from the catchment areas of 200 health facilities. In Tanzania, three-quarters of the 200 communities were randomly selected one each from the catchment areas of health facilities, and one-quarter were randomly selected in groups of two from the catchment areas of other health facilities. See Arkedis et al. (forthcoming) for a comprehensive discussion.

In Ghana and Sierra Leone, communities were randomly selected from the catchment areas of health facilities in each district stratified into three groups according to the type or scale of health facility available to them and their distance to the district capital, where the district's most developed health facilities were located. In Ghana, the three strata included rural communities of at least 400 residents (with two exceptions noted below) in the catchment areas of 1) facilities with a midwife; 2) facilities with no midwife but located closer than the average community to the district capital, where the district hospital was located and where district supply chains of drugs and medical equipment originated; and 3) facilities with no midwife and located farther than the district average from the district capital. From the first two strata, two communities (one each from the catchment areas of two facilities) were randomly selected to be offered the program, and another community from the catchment area of a third facility was randomly selected to be in the control group; from the catchment areas of facilities in the third stratum, one community was randomly selected to be offered the program and one to be part of the control group. In the catchment areas of two district facilities in the control group, all villages had fewer than 400 residents. In these catchment areas, we selected the most populated communities (populations ~300 and ~250). In Sierra Leone, communities of at least 250 residents were randomly selected from the catchment areas of eight facilities in three chiefdoms in the district. Two were Maternal and Child Health Posts—the smallest facilities in Sierra Leone's health system, each serving from 500 to 5,000 people within a 3-mile catchment area—one of which was selected to be part of the control group along with two of the remaining six Community Health Posts, larger facilities serving 5,000–10,000 people within a 5-mile catchment area. A ninth facility was not included because staff of the local partner organization knew that its quality was unusually high for the district, and two communities were not included in either group because the local partner organization had previously worked in them on a program with important similarities to that examined in this paper.

In Malawi, where the program involved traditional authorities, communities of at least 300 residents were selected from three stratified groups of communities in the jurisdictions of Traditional Authorities with either one, two, or three health facilities accessible in the rainy season.[54] From Traditional Authorities with one health facility, one community was randomly selected to be offered the program and one to be part of the control group; from Traditional Authorities with two or three facilities, two communities (one each from the catchment areas of two facilities) were randomly selected to be offered the program, and one from the catchment area of a third facility was randomly selected to be part of the control group.

Appendix B. Two Community Scorecard Programs

This appendix supplements the brief discussion in the paper of the community scorecard programs with additional detail about the information provided to participants in the programs and the positions of the officials who were involved in the program in Ghana, Malawi, and Sierra Leone, as well as a more detailed discussion of the distinction in the theory of change of the program involving officials in Ghana, Malawi, and Sierra Leone from the program in Indonesia and Tanzania and from other community scorecard, transparency and accountability, and social accountability programs.

Statistics and Stories in the Scorecard Meeting

In the first meeting, the facilitator led a discussion of a community "scorecard" (Figure B.1 and Figure B.2). While the scorecard was presented in a different specific format in each country, the overall structure of the scorecard in each community included a brief description of maternal and newborn health-care-seeking among the community and of potential problems with maternal and newborn health care and a statistic or graphical representation of care-seeking or the degree of each problem in the community. The information provided included:

1. maternal mortality rates in the region;
2. results of a survey conducted by the facilitator, in the weeks prior to the meeting, about the prevalence in the community of several health behaviors that medical research has found to be associated with better health, including regular prenatal care and delivering in a facility with a skilled professional; and
3. results from that same survey about 25 potential problems with maternal and newborn health and health care that community members might have experienced and that might be within their capacity to directly or indirectly alleviate, such as the facility being dirty, ill equipped, costly, or difficult to get to, the provider being disrespectful or frequently unavailable, or the prevalence of misunderstandings among community members about effective care-seeking.[55]

Figure B.1. A scorecard from the program in Indonesia

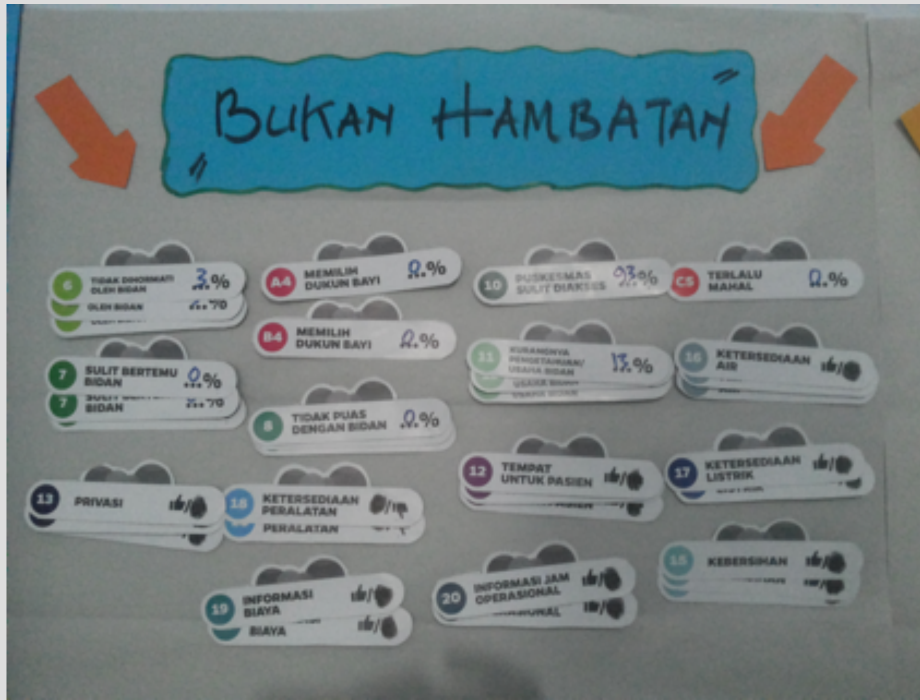


Figure B.2. A scorecard meeting in Tanzania

After this discussion, the facilitator also provided those participating in the discussion with information on approaches that participants might think of trying in order to help to alleviate whatever problems they decide to concentrate on. The facilitator provided this in the form of nine "social action stories," real examples of how nearby communities had fixed similar kinds of problems with their public services, including by:

1. switching to better providers (for those communities with exit options);
2. fixing problems on their own;
3. conversation and negotiation with providers, leading to mutual commitment to solutions;
4. complaining to providers' supervisors;
5. engaging with a "broker" with good connections to government officials;
6. visiting neighboring communities to learn how they had fixed problems;
7. protesting;
8. engaging with media; and
9. publicly recognizing those who are doing a poor job or highlighting those doing an excellent job as highly valued members of the community.

Both community scorecard programs we study were designed by nongovernmental organizations in each country with experience in transparency and accountability, governance, community engagement, and community health as well as by an international team of practitioners and scholars, the authors of this paper among them, to make the workings of each community's maternal and newborn health-care services[56] more transparent and to encourage community members to deliberate on and then try a plan of activities they thought would improve the quality and accessibility of those services. These NGOs were: in Ghana, the Center for Democratic Development; in Indonesia, PATTIRO, a research and policy advocacy organization focused on regional and local governance issues in a number of sectors; the Malawi Economic Justice Network; the Water, Sanitation and Hygiene Network (WASH-net) in Sierra Leone; and the Tanzania country office for the Clinton Health Access Initiative (CHAI). The earlier, large-scale program in Indonesia and Tanzania was developed over the course of two years of iterative co-design and piloting.[57] The program in Ghana, Malawi, and Sierra Leone was developed based on the earlier design in Indonesia and Tanzania but with the addition of meetings with officials. Both programs were also designed to avoid suggesting or incentivizing any particular approach to improving the problems that participants decided to focus on, as well as to avoid offering participants additional resources or authority beyond their existing capabilities as members of their communities and citizens of Ghana, Indonesia, Malawi, Sierra Leone, or Tanzania.

Officials Involved

In Ghana and Sierra Leone, the program involved elected and appointed officials, including several with direct formal authority over the health system. In Ghana, the officials involved included the Presiding Member for the District Assembly, the District Director of Health Services, and the District Planning Officer. The Presiding Member is the ranking member of the local assembly and acts as speaker. The District Director of Health Services supervises all health facilities within the district and is responsible for implementing the Ghana Health Service's policies at the district level. The District Planning Officer prepares the medium-term development plan for the district and monitors all development activities. Of these three officials, only the District Director of Health Services has direct authority over health facilities. In Sierra Leone, the officials involved included the District Development and Planning Officer, the District Health Sister, the District Chief Administrator, and three Local Councillors who represented the five communities who would be offered the program.[58] The District Development and Planning Officer coordinates

all district development activities, including of donors and nongovernmental organizations. The District Health Sister is the second-in-charge of the health system in the district, under the District Medical Officer, and is responsible for supervising and allocating health-facility staff. The Chief Administrator is responsible for all financial and resource management and the day-to-day running of the local council. Each of the three Local Councillors represents a ward of several villages, whom they represent in the local council, including in discussions of district development projects. Councillors work part-time; two of the three councillors involved in the program were also teachers. Of the six officials involved in Sierra Leone, only the District Health Sister has direct authority over health facilities.

Rather than elected or appointed officials, the program in Malawi was designed to broker collaboration primarily with traditional authorities.[59] Malawi's hereditary chieftaincy system has four ranks: Paramount Chiefs, Traditional Authorities, Group Village Heads—traditional leaders responsible to the area's Traditional Authority—and Village Heads.[60] The scorecard program involved the Traditional Authority in the area of three of the five communities where the program was offered and the Group Village Head in the other two communities. The logic of involving traditional authorities was based on their substantial informal influence, with which they might broker connections with others in positions of authority and influence and otherwise support community members seeking to improve their health care. Traditional authorities in Malawi are paid small honoraria by the government, have some formally codified, legal authority, including customary responsibilities around land allocation and adjudicating disputes, often interface with nongovernmental organizations and help them distribute aid, and have influence in community health, including maternal and child health care.[61] In addition, those with the specific rank of Traditional Authority have seats on the District Executive Committee, a district decision-making forum that also includes officials with responsibility for district health care and other public services. Health facilities are also organized around the jurisdictions of Traditional Authorities, who formally oversee each Health Center's Advisory Committee (HCAC). But traditional authorities do not control public funds or taxes; rather their influence over the day-to-day functions of government, including the health system, is mostly informal, indirect, and derived from tradition. When in follow-up interviews after the program ended we asked the traditional authorities who had been involved what authority or influence they have over the health system, all five described the ability to meet with the health facility-in-charge, the district health office, or the District Executive Council.

Distinctions in the Programs

Compared with the earlier scorecard program in Indonesia and Tanzania, the distinction in the theory of change of the program involving officials in Ghana, Malawi, and Sierra Leone has important similarities to contemporary theory in the fields of participatory development and social accountability, including the increasing priority in development scholarship and practice of participatory, community-led assistance appropriate to local political contexts, rather than programs standardized and determined by outsiders (OECD 2008; Andrews, Pritchett, and Woolcock 2017; de Renzio 2006; Halloran 2014; Hickey and Bukenya 2020) and of co-production and co-governance across the state-society divide (Fox 2015; Ackerman 2004; Ostrom 1996; Evans 1996). Bringing together citizens and officials in collaborative problem-solving is similar to the "interface" between community members and providers that Björkman and Svensson (2009) found led to dramatically healthier mothers and children in Uganda in 2004, although the interface meetings in Ghana, Malawi, and Sierra Leone involved upper-level officials in positions of influence in the health system.

The program in Malawi shared myriad similarities with the program in Ghana and Sierra Leone. But in addition to the difference in the roles and positions of officials involved, three smaller differences between the program in Malawi and the program in Ghana and Sierra Leone are also consistent with the

different experiences and reflections of participants in Ghana and Sierra Leone relative to those in Malawi and thus may also have played some role in the differences we describe in Section 5 of the main paper.

First, in both Ghana and Sierra Leone the organizations involved in the program had a history of working in the district. In Malawi, the organization was relatively new to the district.

Second, in Ghana and Sierra Leone, the meetings with officials were scheduled separately from existing forums. In Malawi, the first meetings between community participants and traditional authorities were also separate from existing forums, but the final meeting was incorporated into a regular meeting of the District Executive Committee.

Third, in Malawi the program was relatively accelerated. In several communities, the second and third follow-up meetings were held as little as a week apart, compared with a month in Ghana and Sierra Leone. The shorter time frame may have meant less time for building trust and connection between community participants and traditional authorities or other officials.

Finally, although the programs in Ghana and Sierra Leone were similar, there were also differences between them, including the timing of meetings, the number of officials involved (three in Ghana; six in Sierra Leone), and idiosyncratic differences in the way each organization and group of facilitators approached the program. The similarities in the experiences and reflections of participants in Ghana and Sierra Leone imply that these other differences between the program in each had relatively less influence on the differences in their experiences from participants' in Indonesia, Malawi, and Tanzania.

Appendix C. Planned Activities

This appendix supplements the brief discussion in Section 5 of distinct activities that participants planned to try to improve access to quality maternal and newborn health care in their communities with additional discussion of whom else participants planned to involve and their attempts at these activities over the course of the program.

Table C.1 below shows the distribution across communities of the number of activities participants in the two programs planned, and Table C.2 shows examples of six distinct activities planned by participants.

Table C.1. Number of activities planned

	Community scorecard (large scale)			Community scorecard with officials (small scale)				
	Indonesia	Tanzania	Average*	Ghana	Sierra Leone	Average**	Malawi	Average***
% of communities in which the number of distinct activities that participants planned was								
0	0%	0%	*0%*	0%	0%	*0%*	0%	*0%*
1	0%	4%	*2%*	0%	0%	*0%*	20%	*7%*
2	4%	18%	*11%*	20%	0%	*10%*	20%	*13%*
3	7%	45%	*26%*	0%	0%	*0%*	40%	*13%*
4	15%	23%	*19%*	0%	0%	*0%*	20%	*7%*
5	18%	4%	*11%*	20%	60%	*40%*	0%	*27%*
6	15%	5%	*10%*	20%	20%	*20%*	0%	*13%*
7	14%	1%	*8%*	40%	20%	*30%*	0%	*20%*
8	12%	0%	*6%*	0%	0%	*0%*	0%	*0%*
9	15%	0%	*8%*	0%	0%	*0%*	0%	*0%*
10+	0%	0%	*0%*	0%	0%	*0%*	0%	*0%*

Notes: * Average of the program in Indonesia and Tanzania, across all activities planned by participants in each country. ** Average of the program with officials in Ghana and Sierra Leone, across all activities planned by participants in each country. *** Average of the program with officials in Ghana, Sierra Leone, and Malawi, across all activities planned by participants in each country.

Table C.2. Examples of participants' plans

[Note: To protect anonymity, specific dates and individual and community names have been removed.]

A. To engage the District Health Directorate to request a midwife

Goal: To get a midwife for our facility to help ensure that pregnant women deliver free of complications
Barriers addressed: Lack of midwife at the facility
Measure of success: If we get a resident midwife in our community

List Steps	Responsible Person	What tools, community resources are needed?	Timeline/ Deadline	How is success measured?
Meeting with chiefs and elders to discuss the issue of lack of midwife with them	—	No tools required at this stage	—	If chiefs and elders nominate representatives to go and meet the village nurses
Meeting with our Community nurses to discuss requesting a convenient meeting date with the District Health Directorate	—	No tools required at this stage	—	If the messengers give us feedback with a convenient date for an engagement meeting with the DHD
Representatives of chiefs and elders to have engagement meeting with the DHD to request a midwife or to suggest a trained Traditional Birth Attendant (TBA) to them for consideration	—	Taxi & transfer [fare] to be collected from the community for the messengers	—	If our messengers come back and give us feedback confirming the engagement
Hold a meeting to harvest feedback from representatives of chiefs and elders	—	No tools required at this stage	—	If we get feedback from our messengers and we decide on what action to take next
Have a follow up meeting with the Health Directorate based on the feedback of the 1st meeting with them	—	T&T to be collected from the community for the messengers	—	If we succeed in getting a midwife for our community or if the Health Directorate agrees to work with our trained TBA

B. Awareness campaign

Goal: Sensitizing the community on the importance of getting proper antenatal and postnatal checks and seeking treatment for children with fever
Barriers addressed: Lack of knowledge about check ups, cultural and religious beliefs, fear of going for HIV tests, especially husbands, and laziness
Measure of success: Awareness campaign done

List Steps	Responsible Person	What tools, community resources are needed?	Timeline/ Deadline	How is success measured?
Meeting between the [participants] themselves	—	None	—	Meeting between [participants] has been done
Meeting with the village head to ask for mobilization of the community	—	None	—	Village head asked to help with mobilizing the community
Mobilization of the community	—	None	—	Community mobilized
Awareness meeting	—	Health workers, cultural and religious leader	—	Awareness meeting done

C. To advocate with health workers to undertake outreach sessions in the community

Goal: On or before [date redacted] will meet and set outreach session timeline with members of health center
Barriers addressed: Health workers never undertake outreach sessions in the community
Measure of success: Health workers will undertake outreach sessions in the community

List Steps	Responsible Person	What tools, community resources are needed?	Timeline/ Deadline	How is success measured?
Organize meeting with community members and share findings with them	—	a) Venue b) Participation of community members	—	
Engage health workers at—with our Ward Councillor to agree on timeline for outreach sessions plans	—	Health Center	—	
Follow-up with health workers about undertak-ing outreach sessions	—	Health Center	—	
Educate community members to adhere to timeline for outreach sessions agreed upon by health workers	—	Community Barray [circular meeting place]	On-going	

D. To construct one more labour room

Goal: *Enough labour room space*
Barriers addressed: *Not enough space in the labour room*
Measure of success: *Ensure there is space at the labour room*

List Steps	Responsible Person	What tools, community resources are needed?	Timeline/ Deadline	How is success measured?
Meeting with town chief	—	—	—	
Meeting with the councillor	—	—	—	
Meeting with Section Chief to call a town hall meeting	—	Transportation and venue	—	
[Participants] and town chief to meet and inform in charge about the activity	—	—	—	
[Participants] and Chief to engage councillor and MP to seek financial support and local material	—	—	—	

E. Meeting the ward councillor and development committee members to advocate for the construction of an under-five clinic in the area

Goal: *To advocate for the construction of an under-five clinic in the area*
Barriers addressed: *Long distance to the health center*
Measure of success: *The meeting with the councillor and development committee members done*

List Steps	Responsible Person	What tools, community resources are needed?	Timeline/ Deadline	How is success measured?
Meeting of [participants]	—	N/A	—	Meeting done
Writing letters to the councillor and development committee members	—	Pen and paper	—	Letters delivered
Meeting conducted	—	N/A	—	Meeting done

F. To seek assistance from our MP for the construction of a delivery ward and accommodation for the midwife

Goal: *To acquire a labour ward and accommodation for our midwife*
Barriers addressed: *Lack of midwife at the CHPS [Community-based Health Planning and Services] Compound*
Measure of success: *If we succeed in acquiring a labour ward and accommodation for our midwife*

List Steps	Responsible Person	What tools, community resources are needed?	Timeline/ Deadline	How is success measured?
Meeting with the chief and elders to discuss the issues with them	—	No resources needed	—	If the chief and elders agree to discuss with the MP
Chief and elders to discuss the issue with the MP and fix a day for the meeting	—	No resources needed	—	If the MP agrees to meet the community and fix a convenient date for the engagement
MP to visit the community for community representatives and chief and elders to discuss the issue of constructing a labour ward and accommodation for midwife with him	—	No resources needed	—	If we have engagement with the MP and discuss the issue of constructing the facilities with him
Have follow up calls and meeting with the MP on the issue	—	Airtime for communication to be gotten from the community	—	If the MP gives feedback to our chief and elders to help us construct the labour ward and the accommodation for the midwife
Mobilize materials for the construction of the facilities	—	A storage room for the construction materials to be gotten from the MP, village youth for packing construction materials	—	If we are able to mobilize about 30% of the construction materials and tools
Mobilize the community to carry out communal labour for the construction	—	Construction materials to be gotten from the community, food for the workers to be gotten from the community	—	If the facilities are 100% completely constructed
Hand over the construction of completed facility to the Community Health Committee through a durbar	—	No resources needed	—	If the facility is handed over to the Community Health Committee

Others Participants Planned to Involve

Most of participants' planned activities involved a small number of actors with whom participants primarily planned to engage, with others involved along the way. We first consider the role or position of actors participants planned primarily to involve, and then the roles or positions of any of the actors participants planned to involve.

Roles of primary actors participants planned to involve

The following tables compare the primary actors with whom participants planned to engage in the two programs. Table C.3 expands on the discussion in Section 5 of the main paper by showing the proportion of all activities in which participants planned primarily to engage with others in their community, health-care providers, long-route actors including officials in the village government or in the district government or above, or other actors such as Traditional Birth Attendants.

Table C.3. Primary actors involved in all planned activities

	Community scorecard (large scale)			Community scorecard with officials (small scale)				
	Indonesia	Tanzania	Average*	Ghana	Sierra Leone	Average**	Malawi	Average***
% of activities involving . . .								
Others in the community[1]	50%	76%	*63%*	31%	24%	*28%*	44%	*33%*
Health-care providers (short route)	40%	27%	*34%*	16%	36%	*26%*	50%	*34%*
Long route								
Village government	18%	30%	*24%*	22%	3%	*12%*	11%	*12%*
Outside community	6%	12%	*9%*	63%	52%	*57%*	6%	*40%*
Other (including Traditional Birth Attendants)	5%	4%	*4%*	9%	3%	*6%*	33%	*15%*

Notes: [1] "Others in the community" includes, for example, pregnant women, mothers, husbands, or community elders. * Average of the program in Indonesia and Tanzania, across all activities planned by participants in each country. ** Average of the program with officials in Ghana and Sierra Leone, across all activities planned by participants in each country. *** Average of the program with officials in Ghana, Sierra Leone, and Malawi, across all activities planned by participants in each country.

Table C.4. shows the proportion of communities in which at least one of the activities participants planned primarily involved others in their community, health-care providers, officials in the village government, officials in the district government or above, or other actors (such as Traditional Birth Attendants).

Table C.4. Primary actors participants planned to involve in at least one activity

	Community scorecard (large scale)			Community scorecard with officials (small scale)				
	Indonesia	Tanzania	Average*	Ghana	Sierra Leone	Average**	Malawi	Average***
% of communities where participants planned to engage with . . .								
Others in the community	99%	100%	*100%*	100%	80%	*90%*	100%	*93%*
Health-care providers (short route)	91%	74%	*83%*	80%	80%	*80%*	100%	*87%*
Long route								
Village government	75%	76%	*76%*	80%	20%	*50%*	40%	*47%*
Outside community	29%	43%	*36%*	100%	100%	*100%*	20%	*73%*
Other (including Traditional Birth Attendants)	13%	18%	*16%*	60%	20%	*40%*	80%	*53%*

Notes: * Average of the program in Indonesia and Tanzania, across all communities in each country. ** Average of the program with officials in Ghana and Sierra Leone, across all communities in each country. *** Average of the program with officials in Ghana, Sierra Leone, and Malawi, across all communities in each country.

Roles of all those whom participants planned to involve in their activities
Table C.5 compares, for the two scorecard programs, the proportion of communities in which participants planned at least one activity involving health-care providers, village-level government, or officials outside the community. For the program involving officials, Table C.6 shows the level of these providers and officials as well as the proportion of communities in which participants planned at least one activity involving traditional authorities and providers and civil society.

Table C.5. All actors participants planned to involve in at least one activity

		Long route	
	Health-care providers (short-route)	Village-level government	Officials, district and higher
Indonesia	99%	98%	35%
Tanzania	85%	98%	53%
Average (Indonesia and Tanzania)	*92%*	*98%*	*44%*
Ghana	80%	100%	60%
Sierra Leone	100%	100%	80%
Average (Ghana and Sierra Leone)	*90%*	*100%*	*70%*
Malawi	40%	100%	20%
Average (Ghana, Sierra Leone, and Malawi)	*73%*	*100%*	*53%*

Note: Percentages are the proportions of communities in which participants planned at least one activity that involved others in each role.

Table C.6. Level of all actors participants planned to involve

Level relative to community		Formal government structures		Traditional structures	Civil society
		Health System	Broader Government		
↑			…regional or national political leadership		
	Ghana		20%		
	Sierra Leone		40%		
	Malawi		0%		
		. . . regional or national health officials	. . . regional or national officials (non-health)		. . . regional or national civil society or media
	Ghana	0%	0%		0%
	Sierra Leone	0%	0%		0%
	Malawi	0%	0%		0%
			…district political leaders		
	Ghana		60%		
	Sierra Leone		60%		
	Malawi		0%		
		. . . district-level health officials	. . . district officials (non-health)	. . . traditional leaders (chief, TA, etc.)	. . . district civil-society organizations or media
	Ghana	40%	40%	0%*	0%
	Sierra Leone	0%	20%	60%	0%
	Malawi	0%	20%	60%	0%
			. . . village-level govern-ment (village long route)		
	Ghana		100%		
	Sierra Leone		100%		
	Malawi		100%		
		. . . health-care provid-ers (short route)		. . . providers of tradi-tional care and others in community	. . . local committees, informal leaders, or power brokers
	Ghana	80%		0%	40%
	Sierra Leone	100%		0%	0%
↓	Malawi	40%		0%	100%

Notes: Percentages are the proportions of communities in which participants planned at least one activity that involved others in each role. * In Ghana, village chiefs are included in village-level government (village long route).

Attempted Activities

Table C.7 classifies activities as of the third follow-up meeting at the end of the program as 1) "complete," if participants in the third follow-up meeting described all the steps they had planned as part of the activity as complete (note that "complete" does not necessarily indicate that the activity was successful at achieving participants' goals) or as 2) "ongoing" (not all steps had been completed but the activity was continuing); 3) "stopped" or "not started" (participants never started the activity and had no further plans to do so); or 4) "unclear."

Table C.7. Progress in activities

	Community scorecard (large scale)			Community scorecard with officials (small scale)				
	Indonesia	Tanzania	Average*	Ghana	Sierra Leone	Average**	Malawi	Average***
Complete	53%	65%	*59%*	56%	36%	*46%*	61%	*51%*
Ongoing	31%	24%	*28%*	19%	64%	*42%*	28%	*37%*
Stopped or not started	10%	8%	*9%*	25%	0%	*13%*	11%	*12%*
Unclear	6%	3%	*5%*	0%	0%	*0%*	0%	*0%*

Notes: * Average of the program in Indonesia and Tanzania, across all activities planned by participants in each country. ** Average of the program with officials in Ghana and Sierra Leone, across all activities planned by participants in each country. *** Average of the program with officials in Ghana, Sierra Leone, and Malawi, across all activities planned by participants in each country.

Table C.8 shows the distribution across communities of the number of activities participants described having completed in the third follow-up meeting with the facilitator.

Table C.8. Completed activities across communities

	Community scorecard (large scale)			Community scorecard with officials (small scale)				
	Indonesia	Tanzania	Average*	Ghana	Sierra Leone	Average**	Malawi	Average***
0	9%	2%	*6%*	0%	20%	*10%*	20%	*13%*
1	6%	9%	*8%*	20%	0%	*10%*	0%	*7%*
2	14%	25%	*20%*	20%	40%	*30%*	20%	*27%*
3	11%	44%	*28%*	20%	0%	*10%*	60%	*27%*
4	22%	16%	*19%*	0%	40%	*20%*	0%	*13%*
5	18%	3%	*11%*	20%	0%	*10%*	0%	*7%*
6	9%	1%	*5%*	0%	0%	*0%*	0%	*0%*
7	8%	0%	*4%*	20%	0%	*10%*	0%	*7%*
8	3%	0%	*2%*	0%	0%	*0%*	0%	*0%*

Notes: * Average of the program in Indonesia and Tanzania, across all planned activities completed by participants in each country. ** Average of the program with officials in Ghana and Sierra Leone, across all planned activities completed by participants in each country. *** Average of the program with officials in Ghana, Sierra Leone, and Malawi, across all planned activities completed by participants in each country.

Appendix D. Perceptions of Efficacy

Here we expand on the discussion in the paper of participants' perceptions of their own efficacy at realizing improvements. Prior to the first meeting and after the last, we asked participants this question: "I would now like you to think of improvements of any kind that you would like to make to improve life in your village, for yourself and others—for example, improving garbage collection to keep the village clean, fixing a bad road, organizing a watch group to keep the neighborhood safe, or anything else that you think would improve life in this village for yourself and others. Do you feel that you have the power to help make these kinds of improvements to life in this village, for yourself and others?" Participants' answers to this question reflect most participants beginning the program already highly confident in their capacities to improve their communities; only in Indonesia did a majority describe themselves as skeptical (Table D.1).

Table D.1. Individual perceptions of civic efficacy

	Before first meeting				After last meeting						
	Completely able			Unable	Completely able			Unable	Participants whose perceptions of efficacy . . .		
	4	3	2	1	4	3	2	1	increased	were unchanged	decreased
Attended first and last meetings											
Indonesia	9%	30%	49%	12%	28%	37%	34%	1%	46%	46%	7%
Tanzania	80%	18%	2%	1%	74%	16%	5%	2%	13%	67%	20%
Ghana	79%	21%	0%	0%	79%	18%	3%	0%	13%	72%	15%
Sierra Leone	53%	41%	4%	2%	59%	37%	4%	0%	31%	47%	22%
Malawi	100%	0%	0%	0%	100%	0%	0%	0%	0%	100%	0%
Attended only first or last meeting											
Indonesia	6%	26%	50%	18%	16%	30%	50%	5%			
Tanzania	81%	18%	2%	0%	75%	19%	6%	0%			
Ghana	83%	17%	0%	0%	100%	0%	0%	0%			
Sierra Leone	67%	30%	4%	0%	63%	38%	0%	0%			
Malawi	94%	6%	0%	0%	100%	0%	0%	0%			

Notes: Indonesia and Tanzania: 473 respondents before the first meeting; 306 respondents after the last. Of these, 230 attended both meetings, 243 attended only the first, and 76 attended only the last. Ghana and Sierra Leone: 121 respondents before the first meeting; 113 respondents after the last. Of these, 88 attended both meetings, 33 attended only the first, and 25 attended only the last. Malawi: 84 respondents before the first meeting; 73 respondents after the last. Of these, 52 attended both meetings, 32 attended only the first, and 21 attended only the last.

Participants in Indonesia, where 60% of participants described being skeptical of their civic efficacy before the first program meeting, were much more likely after the program to describe themselves as more confident. Among those who participated in both the first program meeting and the third and

final follow-up meeting, nearly half (46%) described themselves in interviews after the last program meeting as more confident than they had in interviews before the first; only 7% described themselves as less confident than they had before they started.

In the other four countries, where very few participants began the program skeptical of their civic efficacy, very few ended the program skeptical. Among participants in Malawi who also participated in the third and final follow-up meeting, no one changed their answers: all said they were fully confident both before the first program meeting and after the last. In the other three countries, some of those who also participated in the third and final follow-up meeting changed their answers, but forthcoming analyses of these interviews suggest that the changes were smaller than among participants in Indonesia. In particular, some who said they were mostly but not fully confident at the start of the program said later, at the end of the program, that they were fully confident, suggesting that they exceeded relatively lower expectations, and some said the reverse: they told interviewers they were fully confident in their civic efficacy before the first meeting and after the last said they were mostly but not fully confident, suggesting that their relatively higher initial confidence had waned since the first meeting. In Ghana and Tanzania, where initial expectations were higher (79% of participants in Ghana and 80% in Tanzania told interviewers before the first meeting that they were fully confident in their civic efficacy), slightly more seemed after the last meeting to have tempered their initial expectations (15% in Ghana and 20% in Tanzania) than exceeded them (13% in both countries). In Sierra Leone, where initial expectations were relatively lower than in any country except Indonesia, more participants ended the program more confident than they began it (31% vs. 22%), although nowhere near the proportion in Indonesia.[62]

Observations of meetings suggest similar optimism. In Ghana, Malawi, and Sierra Leone, all or almost all participants in early meetings appeared to observers to be optimistic in the first and second program meetings that they would be able to improve their community's health care, and by the final follow-up meeting, participants in most communities appeared to have remained optimistic. This is similar to the program in Indonesia, where observations of early program meetings suggest that despite early skepticism about their individual civic efficacy, most participants were optimistic that the plan they had developed would improve health care. The program in Tanzania was an exception: in the final follow-up meetings in Tanzania, participants in half of communities appeared to be skeptical that they would be able to sustain improvements in their health care.[63] Table D.2 shows, over the course of the first two program meetings and the final follow-up meeting, the proportions of communities in which participants seemed to meeting observers to be skeptical of the efficacy of their efforts.

Table D.2. Skepticism among participants (meeting observations)

Proportion of communities in which	Indonesia	Tanzania	Ghana	Malawi	Sierra Leone
Meeting 1: Scorecard					
All or a majority seemed skeptical at the stories of other communities improving their public services*	10%	5%	—	20%	0%
Meeting 2: Planning					
All or a majority seemed skeptical that the final plan of activities would help alleviate the problems they had chosen to focus on	0%	5%	0%	0%	0%
Meeting 6: 3rd Follow-up					
All or a majority seemed skeptical that their efforts would sustain improvements	12%	53%	0%	20%	0%
Everyone seemed skeptical that their efforts would sustain improvements	0%	43%	0%	20%	0%

Note: * In Ghana, facilitators did not tell the stories of other communities making improvements in three of the five communities. (In neither community where the participant told these stories did a majority of participants hearing them appear to observers to be skeptical that it was possible for communities to make improvements on their own in the ways portrayed in the stories.)

Table D.3 summarizes participants' reflections on their efforts and their experiences of the program in focus groups after it: six months afterward in Sierra Leone, one year afterward in Ghana and Malawi, and 1.5 years afterward in Indonesia and Tanzania.

Table D.3. Reflections on participating

	Indonesia	Tanzania	Ghana	Malawi	Sierra Leone
Proportion of communities in which participants in focus groups . . .	1.5 years after the program		1 year after the program		6 months after
recalled a specific activity that they thought had been successful	93%	100%	80%	100%	60%
recalled a specific activity that they thought had been unsuccessful	39%	41%	60%	60%	100%
said that their activities overall had improved health care in the community	83%	95%	100%	100%	100%
said they were still meeting as a group	23%	26%	60%	40%	100%
said their last meeting was					
within the last month	4%	9%	40%	0%	80%
within the last two months	8%	14%	60%	20%	100%
recalled personal benefits from participating					
>75%	94%	100%	100%	80%	80%
>50%	98%	100%	100%	100%	100%
recalled costs or negative consequences from participating					
>75%	15%	19%	0%	60%	0%
>50%	22%	19%	0%	80%	0%
said that they were glad that they participated	97%	100%	100%	80%	100%

Table D.4 lists general as well as specific, tangible changes that participants across the five countries identified as resulting from the efforts they recalled discussing in the meetings.

Table D.4. Changes participants recalled as a result of their efforts

General, unspecific changes*	Specific, tangible changes
Staff more accessible (excludes staff stays in facility/village)	An ambulance
Services now available 24 hours	A service or arrangement for transportation to the facility
Better equipment or drugs (unspecified)	A new oxygen tank
General community awareness	A new building
Health-worker attitude or performance	New staff
Improved cost/affordability	A new generator
Posyandu (community birth clinic in Indonesia) activity	New or improved road
Improved sanitation/clean water at facility (without specifics)	New or improved rooms at the facility (such as delivery room, inpatient room)
Claimed outcome: such as more women giving birth in the facility	New information board or specific information that is now included on an existing board (cost, midwife phone number)
Household toilets	No expired medicine at facility
Cooperation with facility	A new toilet
Collected data	A service or arrangement for cleaning the facility
Vague improvement in "access"	New pump or other new access to clean water
Vague renovations or improved building	A new pharmacy
Partial preparation for electrification	A new or improved maternity waiting home
New outreach services	Electricity
	A new *posyandu* (Indonesia)
	Distribution of birth-preparedness stickers
	New beds
	A new registration counter
	New waste bins
	A new suggestion box
	Staff who now stay in or near the facility
	Facility health-insurance accreditation

Notes: All changes that participants mentioned resulting from their efforts were first classified into the categories listed above and then by whether they described specific, concrete, tangible changes or more general or vague improvements. * General or unspecific changes include participants describing the road as "improved" or the ambulance as "available," for example, rather than remembering a new ambulance, arrangement for cleaning the facility, new staff, or improvements to the road to the facility.

Finally, the proportions of communities described in the main paper as communities where participants eventually realized at least some of the objectives of their activities are based on a combination of interviews with participants and others in their communities about what they did and what they saw their activities lead to, observations about the outcomes of these activities, including by coauthors of the paper, and focus groups with participants after the program.

In the program in Ghana, Malawi, and Sierra Leone, interviews with participants and others in their communities with observations by coauthors of this paper of program meetings and of outcomes described by participants and others suggest that participants in all communities in Ghana and Malawi and 80% in Sierra Leone were able to achieve the objectives of at least some of their activities. In 60% of communities in Ghana, 20% of communities in Malawi, and 80% of communities in Sierra Leone, we traced participants' activities to tangible, observable changes in their community's health care or their access to health care.

In Figure 5 in the main paper, we compare these proportions to the proportions of communities in Indonesia and Tanzania where the first program was offered. These comparisons are to the proportion of communities where interviews with participants and others in their communities shortly after the final program meeting in 65 communities (41 in Indonesia; 24 in Tanzania) and evidence from ethnographic studies in eight communities (4 in Indonesia; 4 in Tanzania) suggest that participants achieved the objectives of some of their activities. In 44% of these communities in Indonesia and 46% in Tanzania, evidence from interviews with those who attended meetings and with others in the community also suggested that at least one activity unrelated to educating or encouraging neighbors to seek maternal and newborn health care in a health facility was successful at achieving participants' objective. In approximately one-third of communities (41 in Indonesia and 30 in Tanzania), participants were able to identify changes in their health care or access to it among their communities resulting from their efforts that were specific and tangible, and that may have made a difference to their access to quality maternal and newborn health care: new or improved buildings, staff, and ambulances; information boards and birth-preparedness materials that may have made facility services easier to access; waste bins, toilets, and sources of clean water; staff living at or near the facility; and many more (Table D.4). In the 65 communities in which interviewers spoke with those who attended meetings and with others in the community, they heard about similar changes in 26 (40%). In 17 (26%), similar changes were also observed in facilities or described in interviews with recently pregnant women 1.5 years after the program.

Appendix E. Responses by Officials

Table E.1 shows the number of responses that officials in Ghana and Sierra Leone recalled in open-ended interviews, the number of these that we could verify by triangulating with other interviews, focus groups, and observations, and the number that were more broadly encompassing reforms with effects beyond the participants' community or facility.

Table E.1. Verified responses

	Ghana	Sierra Leone
Specific responses described in open-ended interviews	9	6
Strong (weak) triangulating evidence of the response	3 (1)	2 (1)
Encompassing responses	0	0

Here we expand the discussion in Section 5 of responses that we could verify by triangulating with other interviews, focus groups, and observations.

In Ghana, participants worked with an official to open a new facility that had sat unopened for three years because beds had not been delivered. Participants planned to request beds from higher-level officials but learned at the first government interface meeting that the beds had been acquired: they just had not been moved into the facility. One of the officials involved subsequently wrote this request into the district's development plan. In parallel, participants approached other officials to ask them to move the beds and open the facility, which they did. The new facility is larger and community members report it has a better supply of drugs than the previous facility.

Second, an official worked to allocate funds for a new, up-to-standard health facility. This action stemmed from a feud between two villages that shared a health facility. The health facility was physically located in the other village (not part of the community scorecard program), so participants initially planned to try to build a facility in their village. When they presented their plan at the first government interface meeting, the official informed them that they could not build a separate facility in their village because it would be less than the standard five miles away from the existing health facility. However, the existing health facility was also not up to standard, so the official instead suggested that participants work together with residents of the other village to request and identify land for an upgraded health facility between the two villages. Following the meeting, participants worked with the elders in their community to resolve the feud with the other village and identified land for the facility, cleared the land, and worked with the official to request funding for the updated facility, which was granted. Subsequent observations showed that construction had started on the new facility.

Third, an official arranged for representatives of the National Insurance Scheme to explain fees and charges and repaired relations between facility staff and community members concerned about fees they were paying. Members of one community involved in the scorecard program were unhappy with a staff member at their health facility being disrespectful and charging fees for services that they understood to be covered by insurance. As a first step, participants invited members of the national health-insurance scheme to explain their entitlements to the community. Two months later, participants arranged a meeting with the official to lodge their complaints about the staff member. The official subsequently arranged for the staff member to be replaced and transferred to a larger health facility where they would be under closer supervision. Community members subsequently reported that they were happy with the staff's replacement.

In Sierra Leone, one official connected participants with a grant writer in the district, who helped them secure support from a nongovernmental organization to repair the pump at their facility and dig a second well to improve the facility's supply of water. Participants initially planned to have the water pump at the local health facility repaired. They took this request to the official, who brought it to the District Council, but they were unable to fund it, so the official subsequently connected participants to a grant writer who prepared proposals to several NGOs for the well and other requests (toilets as well as community meeting space). One of the NGOs responded to the request and not only repaired the current pump but dug a second well.

A second official helped to mediate a conflict between a community and facility staff that led to a substantial although short-term improvement in their relationship and in community members' perceptions of how they were treated at the facility. The participants planned to try to improve the relationship between their community and the facility staff, whom they perceived to be disrespectful, to charge for services that should be free, and to be unwilling to attend to them during emergency situations after hours. They shared their plan with the official, who subsequently coordinated a meeting between them and the health-facility staff. During the meeting, which the official mediated, the facility staff promised to be more welcoming and respectful of patients. In the short term, participants as well as other community members reported an improvement in the way they were treated at the health facility, although the improvement did not last: over time the relationship seems to have deteriorated.

Appendix F. Anchoring Vignettes

In interviews prior to the start of the program and again at the end, we asked participants in the program involving officials which of four anchoring vignettes (Masset 2015; King et al. 2004) was most familiar: one in which officials had listened and responded by trying to improve a public service and three others in which officials were remote or unable to help. Table F.1 shows proportions of responses summarized in Figure 6 in the main paper.

Table F.1. Perceptions of responsiveness

	Proportion of participants for whom the most familiar story . . .				Proportion of participants whose responses . . .		
	. . . before the first meeting, was . . .		. . . after the last meeting, was . . .				
	. . . the story of official listening and responding	. . . one of the three stories of other responses	. . . the story of official listening and responding	. . . one of the three stories of other responses	changed to the story of the most responsive official	remained unchanged	changed from the story of the most responsive official to any of the other three stories
Attended first and last meetings							
Ghana	54%	46%	67%	33%	21%	72%	8%
Sierra Leone	14%	86%	69%	31%	59%	37%	4%
Malawi	24%	76%	33%	67%	25%	58%	16%
Attended only first or last meeting							
Ghana	67%	33%	0%	100%			
Sierra Leone	7%	93%	48%	52%			
Malawi	31%	69%	63%	38%			

Notes: Total respondents who attended both meetings: 39 in Ghana, 55 in Malawi, and 49 in Sierra Leone. Total respondents who attended only the first meeting: 6 in Ghana, 32 in Malawi, and 27 in Sierra Leone. Total respondents who attended only the last meeting: 1 in Ghana, 21 in Malawi, and 24 in Sierra Leone.

The complete vignettes are as follows:
I'm going to tell you about a situation in a village that is similar to yours but has a problem with its school. Sometimes the teacher in this school does not show up to teach, and even when the teacher does come he often does not teach; he's at school but leaves the classroom to go and hang out with other teachers. Many parents in the village are concerned that their children are not learning when they go to school, and they think that the school would improve a lot if the teacher would come more regularly and work harder to teach the students well.

Vignette 1: *Three parents are very frustrated by the poor quality of their village's school. One day they go to the teacher to discuss the problems, but the teacher is not helpful. He denies that he misses class and avoids teaching*

when he's there. After talking to the village head about the problem, the three parents decide to ask the district education director for help. They contact him and obtain a meeting time. At the meeting the district education director listens carefully to their stories and requests, and promises that he will do what he can. The parents leave feeling confident, but over the coming weeks and months they do not hear anything more from the district education director and the teacher still misses class and avoids working hard to teach the students well.

Vignette 2: *Three parents are very frustrated by the poor quality of their village's school. One day they go to the teacher to discuss the problems, but the teacher is not helpful. He denies that he misses class and avoids teaching when he's there. After talking to the village head about the problem, the three parents decide to ask the district education director for help. They try to contact him and obtain a meeting time, but are unsuccessful: every time they phone his office or try to visit, he is always away, in other meetings, or too busy to meet them. The parents are unsure what to do next, and meanwhile the teacher still misses class and avoids working hard to teach the students well.*

Vignette 3: *Three parents are very frustrated by the poor quality of their village's school. One day they go to the teacher to discuss the problems, but the teacher is not helpful. He denies that he misses class and avoids teaching when he's there. After talking to the village head about the problem, the three parents decide to ask the district education director for help. They contact him and obtain a meeting time. At the meeting the district education director listens carefully to their stories and requests, but he tells them that he is unable to help: he explains that teachers are assigned and managed by other officials in the government, not him. The parents are unsure what to do next, and meanwhile the teacher still misses class and avoids working hard to teach the students well.*

Vignette 4: *Three parents are very frustrated by the poor quality of their village's school. One day they go to the teacher to discuss the problems, but the teacher is not helpful. He denies that he misses class and avoids teaching when he's there. After talking to the village head about the problem, the three parents decide to ask the district education director for help. They contact him and obtain a meeting time. At the meeting the district education director listens carefully to their stories and requests, and promises that he will do what he can. A few weeks later, the parents learn that the district education director met with the teacher, and they notice that the teacher starts coming to class regularly, working harder to teach the students well.*

After each vignette, interviewers asked participants to rate, on a 1–4 scale, how responsive the district education director was (meaning how well he listened and tried his best to respond). At the end of the survey, interviewers also asked which story was most familiar.

Appendix References

Acemoglu, Daron, Tristan Reed, and James A. Robinson. 2014. "Chiefs: Economic Development and Elite Control of Civil Society in Sierra Leone." *Journal of Political Economy* 122 (2): 319–68. https://doi.org/10.1086/674988.

Ackerman, John. 2004. "Co-Governance for Accountability: Beyond 'Exit' and 'Voice.'" *World Development* 32 (3): 447–63. https://doi.org/10.1016/j.worlddev.2003.06.015.

Andrews, Matt, Lant Pritchett, and Michael Woolcock. 2017. *Building State Capability: Evidence, Analysis, Action*. New York: Oxford University Press. https://bsc.cid.harvard.edu/building-state-capability-evidence-analysis-action.

Arkedis, Jean, Jessica Creighton, Akshay Dixit, Archon Fung, Stephen Kosack, Dan Levy, and Courtney Tolmie. Forthcoming. "Can Transparency and Accountability Programs Improve Health? Experimental Evidence from Indonesia and Tanzania." *World Development*.

Basurto, Maria Pia, Pascaline Dupas, and Jonathan Robinson. 2020. "Decentralization and Efficiency of Subsidy Targeting: Evidence from Chiefs in Rural Malawi." *Journal of Public Economics* 185: 104047. https://doi.org/10.1016/j.jpubeco.2019.07.006.

Björkman, Martina, and Jakob Svensson. 2009. "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda." *The Quarterly Journal of Economics* 124 (2): 735–69. https://doi.org/10.1162/qjec.2009.124.2.735.

Björkman Nyqvist, Martina, Damien de Walque, and Jakob Svensson. 2017. "Experimental Evidence on the Long-Run Impact of Community-Based Monitoring." *American Economic Journal: Applied Economics* 9 (1): 33–69. https://doi.org/10.1257/app.20150027.

Butrick, E., N. Diamond-Smith, N. Beyeler, D. Montagu, and M. Sudhinaraset. 2014. "Strategies to Increase Health Facility Deliveries: Three Case Studies." San Francisco: University of California, Global Health Group, Global Health Sciences. https://globalhealthsciences.ucsf.edu/sites/globalhealthsciences.ucsf.edu/files/pub/pshi-strategies-to-increase-health-facility-deliveries.pdf.

Carlson, Elizabeth, and Brigitte Seim. 2020. "Honor among Chiefs: An Experiment on Monitoring and Diversion Among Traditional Leaders in Malawi." *Journal of Development Studies* 56 (8): 1541–57. https://doi.org/10.1080/00220388.2019.1703955.

Chansa, Collins, and Anooj Pattnaik. 2018. "Expanding Health Care Provision in a Low-Income Country: The Experience of Malawi." Washington, DC: World Bank. http://documents.worldbank.org/curated/en/355601516177706823/Expanding-health-care-provision-in-a-low-income-country-the-experience-of-Malawi.

de Renzio, Paolo. 2006. "The Primacy of Domestic Politics and the Dilemmas of Aid: What Can Donors Do in Ethiopia and Uganda?" Overseas Development Institute. https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/748.pdf.

Eckstein, Harry, and Ted Robert Gurr. 1975. *Patterns of Authority: A Structural Basis for Political Inquiry*. New York: John Wiley & Sons.

Erlangga, Darius, Shehzad Ali, and Karen Bloor. 2019. "The Impact of Public Health Insurance on Healthcare Utilisation in Indonesia: Evidence from Panel Data." *International Journal of Public Health* 64 (4): 603–13. https://doi.org/10.1007/s00038-019-01215-2.

Escribano-Ferrer, Blanca, Francoise Cluzeau, Derek Cutler, Christiana Akufo, and Kalipso Chalkidou. 2016. "Quality of Health Care in Ghana: Mapping of Interventions and the Way Forward." *Ghana Medical Journal* 50 (4): 238–47. https://doi.org/10.4314/gmj.v50i4.7.

Evans, Peter. 1996. "Government Action, Social Capital and Development: Reviewing the Evidence on Synergy." *World Development* 24 (6): 1119–32. https://doi.org/10.1016/0305-750X(96)00021-6.

Fox, Jonathan. 2015. "Social Accountability: What Does the Evidence Really Say?" *World Development* 72: 346–361. https://doi.org/10.1080/09614520701469955.

Ghana Health Service. 2018. "The Health Sector in Ghana: Facts and Figures." Accra, Ghana: Ghana Health Service. http://ghanahealthservice.org/downloads/Facts+Figures_2018.pdf.

Halloran, Brendan. 2014. "Thinking and Working Politically in the Transparency and Accountability Field." Transparency and Accountability Initiative. https://www.transparency-initiative.org/wp-content/uploads/2017/03/thinking-and-working-politically-may-2014.pdf.

Hickey, Sam, and Badru Bukenya. 2020. "The Politics of Promoting Social Cash Transfers in Uganda: The Potential and Pitfalls of 'Thinking and Working Politically.'" *Development Policy Review* 2020 (00): 1–20. https://doi.org/10.1111/dpr.12474.

King, Gary, Christopher J. L. Murray, Joshua A. Salomon, and Ajay Tandon. 2004. "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research." *American Political Science Review* 98 (1): 191–207. https://doi.org/10.1017/S000305540400108X.

Kosack, Stephen, and Archon Fung. 2014. "Does Transparency Improve Governance?" *Annual Review of Political Science* 17: 65–87. https://doi.org/10.1146/annurev-polisci-032210-144356.

Kosack, Stephen, Grant Bridgman, Jessica Creighton, Courtney Tolmie, and Archon Fung. 2019. "Encouraged Participation." Working Paper. https://ash.harvard.edu/files/ash/files/encouraging_participation.pdf.

Levitsky, Steven, and Lucan A. Way. 2010. *Competitive Authoritarianism: Hybrid Regimes After the Cold War*. New York: Cambridge University Press.

Maharani, Chatila, Desie Frihandini Afief, Dorothea Weber, Michael Marx, and Svetla Loukanova. 2019. "Primary Care Physicians' Satisfaction after Health Care Reform: A Cross-Sectional Study from Two Cities in Central Java, Indonesia." *BMC Health Services Research* 19 (1): 290. https://doi.org/10.1186/s12913-019-4121-2.

Marshall, Monty G., and Ted Robert Gurr. 2014. "Polity IV Project." Center for Systemic Peace. https://www.systemicpeace.org/polity/polity4.htm.

Mboi, Nafsiah. 2015. "Indonesia: On the Way to Universal Health Care." *Health Systems and Reform* 1 (2): 91–97. https://doi.org/10.1080/23288604.2015.1020642.

OECD. 2008. *Paris Declaration on Aid Effectiveness (2005); Accra Agenda for Action (2008)*. Paris: OECD. http://www.oecd.org/dac/effectiveness/34428351.pdf.

Ostrom, Elinor. 1996. "Crossing the Great Divide: Coproduction, Synergy, and Development." *World Development* 24 (6): 1073–87. https://doi.org/10.1016/0305-750X(96)00023-X.

Raffler, Pia, Daniel N. Posner, and Doug Parkerson. 2019. "The Weakness of Bottom-Up Accountability: Experimental Evidence from the Ugandan Health Sector." Working Paper. https://www.poverty-action.org/sites/default/files/publications/The%20Weakness%20of%20Bottom-Up%20Accountability.pdf.

Transparency for Development. 2017. "Citizen Voices, Community Solutions: Designing Better Transparency and Accountability Approaches to Improve Health." Ash Center Occasional Papers. Cambridge, MA: Ash Center for Democratic Governance and Innovation, Harvard Kennedy School. https://ash.harvard.edu/files/ash/files/citizen_voices_community_solutions.pdf.

UNDP. 1990. *Human Development Report 1990: Concept and Measurement of Human Development*. New York: Oxford University Press for the United Nations Development Program. http://www.hdr.undp.org/en/reports/global/hdr1990.

UN General Assembly. 2000. *United Nations Millennium Declaration, Resolution Adopted by the General Assembly*. Vol. A/RES/55/2. http://www.refworld.org/docid/3b00f4ea3.html.

———. 2015. *Transforming Our World: The 2030 Agenda for Sustainable Development*. Vol. A/RES/70/1. http://www.refworld.org/docid/57b6e3e44.html.

UNICEF. 2018. "Tanzania Health Budget Brief 2018." New York: UNICEF. https://www.unicef.org/esa/media/2331/file/UNICEF-Tanzania-Mainland-2018-Health-Budget-Brief-revised.pdf.

Walsh, Aisling, Anne Matthews, Lucinda Manda-Taylor, Ruairi Brugha, Daniel Mwale, Tamara Phiri, and Elaine Byrne. 2018. "The Role of the Traditional Leader in Implementing Maternal, Newborn and Child Health Policy in Malawi." *Health Policy and Planning* 33 (8): 879–87. https://doi.org/10.1093/heapol/czy059.

WHO. 2017. "Sierra Leone Annual Report: A Year in Focus 2017." Geneva: World Health Organization. https://www.afro.who.int/sites/default/files/2018-03/World%20Health%20Organization%20Sierra%20Leone%20Annual%20Report%202017.pdf.

World Bank. 2020. World Development Indicators. https://databank.worldbank.org/source/world-development-indicators.

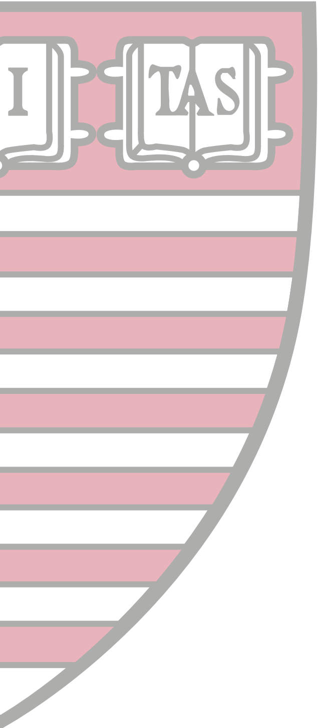
Notes

1. One example of the growth of such programs is the UK Department for International Development's "empowerment and accountability" portfolio, which at the time of a 2016 review had 2,379 projects, including 180 in its 28 priority countries that used a range of approaches to encouraging social accountability (Holland and Schatz 2016, 16).
2. See in particular the Human Development Reports (UNDP 1990 and subsequent reports); the Millennium and Sustainable Development Goals (UN General Assembly 2000, 2015); and the Paris and Accra Declarations (OECD 2008). See also Sachs (2005); Sen (1999); and World Bank (2004).
3. Among others, Banerjee, Banerji, et al. (2006); Banerjee, Hanna, et al. (2018); Björkman and Svensson (2009); Björkman Nyqvist, de Walque, and Svensson (2017); Holland and Schatz (2016); and World Bank (2018). The growing number of reviews of the effectiveness of transparency and accountability and social accountability programs includes 3ie (2018); Fox (2007a, 2015); Joshi (2008, 2010); Joshi and Houtzager (2012); J-PAL (2011); Kosack and Fung (2014); Lieberman, Posner, and Tsai (2014); and McGee and Gaventa (2011). Cornwall (2002, 2004) and Gaventa (2006) review and situate broader spaces of institutionalized and informal participation.
4. See, among others, Björkman and Svensson (2009); Björkman Nyqvist, de Walque, and Svensson (2017); Gullo, Galavotti, and Altman (2016); and Raffler, Posner, and Parkerson (2019).
5. For example, among eight community scorecard programs in five countries supported by CARE reviewed in Gullo, Galavotti, and Altman (2016), three involved district officials; one of these also involved national officials.
6. Among others, see Ackerman (2004); Cornwall (2004); Fox (2015, 2007a); Joshi (2008); and Joshi and Houtzager (2012). Earlier foundational work on co-production and co-governance more generally includes, among others, Evans (1996) and Ostrom (1996).
7. Among others, Chambers (1997); Fishkin and Luskin (2005); Fung (2006a); and Fung and Wright (2001, 2003).
8. See Banerjee et al. (2010); Jenkins and Goetz (1999); Olken (2007); and related discussions in Cornwall (2002, 2004); Fox (2007b, 2015); Gaventa (2006); Joshi and Houtzager (2012); Kosack and Fung (2014); Lieberman, Posner, and Tsai (2014); and McGee and Gaventa (2011).
9. See also Hernández et al. (2019) and Olken (2007).
10. See also Mainwaring and Welna (2003). Fox (2015) describes this form of large-scale, enduring society-state collaboration "acting to offset anti-accountability forces that are often also linked across the state-society divide" and that are "deeply embedded in both state and society [and] are often stronger than pro-accountability forces," making "resistance likely and conflict … both expected and necessary" and creating the need for scaled-up collective voice and action that can exert pressure from below.
11. The other three contextual "worlds" in the framework are those where competition among providers offers citizens exit options (world 1) and where providers do (world 2) and do not (world 3) have incentives and willingness to engage in short-route reform (Kosack and Fung 2014).
12. When coining the "voice" concept in his theory of how people might improve organizational underperformance, Hirschman (1970) was writing about democratic settings.
13. For example, Baiocchi, Heller, and Silva (2011); Fung (2006a); Gaventa (2006); Mansuri and Rao (2013); and Olson (1965). Kahneman and Tversky (1979) and Tversky and Kahneman (1991) find a general tendency to ignore or discount capacities for improvement relative, for example, to efforts to avoid decline in the status quo.
14. See also Ansell and Gash (2008); Donahue and Zeckhauser (2012); and Fung (2006b).
15. Cornwall (2002, 20) argues that these kinds of spaces can provide "those usually excluded from public policy processes with opportunities to lever open policy and political space, and to forge links with other spaces through which networks and alliances can form around particular options or positions," but that they ultimately play a relatively insignificant role in the policy process and may be no more than "an elaborated form of market research, testing out and assessing potential reactions to new policy products, and informing ways of marketing them better, but not fundamentally changing anything much at all."
16. Putnam, Leonardi, and Nanetti (1994) find that the trust and social capital relevant to effective democratic governance can take generations to build. Many scholars have noted growing trust deficits in state-society relations as the third wave of democratic opening has slowed; see, among many, Walker (2016); Armah-Attoh, Selormey, and Houessou (2016); Bratton and Gyimah-Boadi (2016); Brixi, Lust, and Woolcock (2015); and Pew Research Center (2017).
17. Authors of these papers and discussants highlighted the potential importance of this difference during a roundtable discussion at the 2019 meetings of the American Political Science Association where these papers were presented (Björkman Nyqvist et al. 2019).

18. See Arkedis et al. (forthcoming) and Kosack et al. (2019).
19. On commitments and progress both made and remaining to be made in universal access to quality health care, see, among others, Farmer et al. (2013); Hsia et al. (2012); Kruk et al. (2016, 2018); Sen (1999); UNDP (1990); UN General Assembly (2000, 2015); World Bank (2004); and World Economic Forum (2015).
20. In Przeworski et al.'s (2000) definition, multiple peaceful transfers of national political power through regularly contested democratic elections are the primary observable implication of democratic institutions designed to create responsiveness and accountability to citizens' preferences.
21. Tanzania's Polity Index score increased to +3 from −1 the year before the program. More recently this trend has reversed (Freedom House 2020).
22. The resulting programs were similar in many respects to the standard community scorecard program, but not all: in particular, neither included an interface meeting between community members and frontline health-care providers. See Appendix B; Arkedis et al. (forthcoming); and Transparency for Development (2017).
23. The second program had a slightly expanded focus on maternal, newborn, and child health (MNCH rather than MNH).
24. Facilitators were trained to invite a broadly representative group of community members, particularly non-elite (Transparency for Development 2017). Ethnographic studies in eight of the communities by Iqra Anugrah, Megan Cogburn, Mohamed Yunus Rafiq, and Kankan Xie suggest that many also relied on advice from village leadership, and some relied exclusively on village leaderships' suggestions.
25. For more on the design of the program see Appendix B and Transparency for Development (2017); on the contexts where it was offered see Appendix A and Transparency for Development (2016).
26. See Appendix B for a more detailed discussion of the distinction in the theory of change of this program from other community scorecard, transparency and accountability, and social accountability programs.
27. This information on health policy and governance was discussed in the second meeting, in which participants deliberated on how they might alleviate the problems they had decided to focus on in the first meeting. Facilitators varied in their emphasis on this information; observations of meetings suggest that it was rarely a primary focus of meeting discussions and in several communities was not discussed at all.
28. The two where the organizations had worked were both in Sierra Leone.
29. Although this program was similar in all three countries, there were other, smaller differences in the program in each, several of which are also consistent with differences in the experiences of participants, albeit not as directly. These include differences in whether the interface meetings were scheduled separately from existing forums such as regular meetings of governmental committees, the length of time of the program, and the experience and familiarity of the nongovernmental organization with the district. See Appendix B for a more detailed discussion.
30. The program initially also involved the District Health Sister, who would have been the only one of the six officials involved in Sierra Leone with direct authority over health facilities but who did not end up attending either of the two interface meetings.
31. Neither Indonesia nor Tanzania has similar structures of hereditary authority.
32. There are four ranks of traditional authorities in Malawi: Paramount Chiefs, Traditional Authorities, Group Village Heads—chiefs responsible to the area's Traditional Authority—and Village Heads. Village Heads are also called village chiefs. In 2017, there were 3,994 Group Village Heads and 264 Traditional Authorities in Malawi (Carlson and Seim 2020). Basurto, Dupas, and Robinson (2020) study a program involving chiefs in two districts in the same region of Malawi as the communities where the program was offered and find that Group Village Heads were responsible for between 2 and 10 villages.
33. Our approach draws from the large methodological social science and international development literature considering the characteristics of rigorous and reliable observations and their integration in evaluating key observable implications of theory, including Lijphart (1971); Collier (1993); King, Keohane, and Verba (1994); Snyder (2001); Lieberman (2005); Brady (2008); Mahoney (2008); Woolcock (2013); Seawright (2016); Stern et al. (2012); and Giraudy, Moncada, and Snyder (2019). See Transparency for Development (2015) for a more detailed methodological discussion. Interview and observation protocols are available at t4d.ash.harvard.edu.
34. At several points we also rely on interviews and facility observations in all communities 1.5 years after the program from Arkedis et al. (forthcoming) and ethnographic studies in eight of the communities in Indonesia and Tanzania where the large-scale program was offered.
35. See also Creighton et al. (2020) for discussion and additional analyses.

36. We thank Anuradha Joshi for suggesting the use of vignettes for more valid comparisons of participants' perceptions of civic efficacy. Masset (2015) uses similar vignettes to examine perceptions of empowerment or self-efficacy: individuals' belief in their capabilities (Bandura 1977, 1982).
37. Aside from one official who was no longer in a post in the district, interviews included all officials involved regardless of the degree or intensity of their involvement.
38. See Appendix A for further detail on random selection. To increase representativeness of interviews and observations in Indonesia and Tanzania, communities where meetings were observed and participants interviewed were selected randomly from the same national and regional stratifications as those that were offered the program as part of the broader randomized controlled trial.
39. See Appendix C for a discussion of participants' activities and examples of plans.
40. See Creighton et al. (2020) for a discussion of the categories.
41. Table 6 suggests several other differences between participants' plans in the two programs. First, in both programs, participants in between half and two-thirds of communities planned efforts to improve patients' experiences, but in 42% of communities in Indonesia and 36% in Tanzania participants planned to try to improve the availability of information about facility costs or opening hours or install suggestion or complaint boxes; none of participants' plans in Ghana or Sierra Leone included these objectives. Second, in Ghana, participants' plans were also slightly more likely to include efforts to improve their community's access to health care or their relationships with health-care workers (60% of communities in Ghana compared with 41% in Indonesia, 31% in Tanzania, and 40% in Sierra Leone). Third, in Sierra Leone, participants in only 40% of communities planned demand-side efforts. Fourth, although participants in Ghana focused as often on demand-side issues as participants in Indonesia and Tanzania (100% of communities in Ghana, 96% in Indonesia, and 99% in Tanzania), their approaches differed: they were less focused on efforts to increase awareness within their communities of the importance of planning for and seeking care during birth (60% of communities in Ghana, compared with 91% in Indonesia and 95% in Tanzania) or pursuing bylaws, partnerships, or other ways of increasing use of the facility (no communities in Ghana, compared with 16% in Indonesia and 54% in Tanzania). In Ghana it was also slightly more common for participants to focus on community members' ability to pay for care (60% in Ghana compared with 44% in Indonesia and 46% in Tanzania). Finally, participants in both Ghana and Sierra Leone were more likely to plan to seek improvements outside of health care. In Indonesia and Tanzania, participants' plans focused exclusively on health care; in Ghana and Sierra Leone, participants' plans included non-health-related efforts in 40% of communities.
42. In Ghana, all three officials described in Section 3 above who at the start of the program had expressed willingness to be involved in helping participants to improve their maternal and newborn health care attended the meetings. In Sierra Leone, four of the six attended.
43. For example, one official clarified confusion among participants over the payments health facilities could request. Officials typically also engaged in lengthy discussions both among themselves and together with representatives of nongovernmental organizations.
44. For example, one official offered his phone number, and interviews with the official as well as community participants indicate that several subsequently contacted him.
45. This is not an estimate of the proportion of responses that actually happened as described; responses might be unverified because the sources of data described in Section 4 did not include questions with which to triangulate them.
46. As noted above when discussing responses in Ghana and Sierra Leone, this triangulation helps verify and better understand this response but does not provide an estimate of the proportion of responses that actually happened as described, as responses might be unverified because the sources of data described in Section 4 did not include questions we could use to triangulate them.
47. Despite their far more limited engagement with government officials over the course of the program, interviews with participants in Malawi before and after the program also suggest that responsive officials had become more familiar among 25% of participants, similar to the proportion in Ghana (21%) although far less than among participants in Sierra Leone (59%); see Figure 6.
48. This is the only such risk of which we are aware in any of the communities in all five countries, but risks like this have been described as a concern by some in the community of practice around transparency and accountability programs—see, for example, Thampi (2017).
49. This pattern also supports work on the importance of state capacity—e.g., Andrews, Pritchett, and Woolcock (2017); Besley and Persson (2010); Fox (2015); and Grindle (2007, 1997).

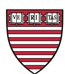
50. See Appendix B, which also notes several other differences between the program in Ghana and in Sierra Leone that had little association with differences in the experiences of participants, including the timing of meetings, the number of officials involved, and idiosyncratic differences in the way organizations and facilitators approached each program.
51. These decades-long commitments are not unusual: much of the world has seen steady improvements in health care following decades of international commitment to sustainable development in part by improving access to quality health care, particularly maternal and newborn health care. See in particular the Human Development Reports (UNDP 1990 and subsequent reports); the Millennium and Sustainable Development Goals (UN General Assembly 2000, 2015); and the Paris and Accra Declarations (OECD 2008).
52. See Fox (2015) and Kosack and Fung (2014).
53. At the time of the program Tanzania's political system was "competitive authoritarian" (+3), defined as those with Polity scores of between −5 and +5 (Levitsky and Way 2010), with a score in the bottom third of countries classified by Polity. In contrast to the selection of three countries with similarities in institutional incentives for vertical democratic accountability, the variation in these institutional incentives between Indonesia and Tanzania in the earlier program allowed inquiry into similarities in the effects of a program focused only in communities across varied political contexts (Arkedis et al., forthcoming).
54. Throughout, "Traditional Authority" refers to the specific rank within Malawi's broader system of traditional authorities. See Section 3 of the main paper as well as Appendix B below.
55. The specific potential problems are: the service provider is difficult to see or not available; the service provider is disrespectful; expecting and recent mothers fear that the medical care offered is unsafe; care is too costly; the service provider lacks skills or knowledge; transportation to the facility is difficult; the facility is difficult to access; the space for patients in the facility is poor quality; there is no privacy in the facility; the facility lacks medicine or supplies; the facility is dirty; there is no water available in the facility; there is no toilet in the facility; the facility lacks a placenta pit or other place where the placenta can be disposed of; there is no reliable electricity or refrigeration in the facility; the facility has poor or missing equipment; expecting and recent mothers are dissatisfied with the service provider; lack of information on the cost of treatment; lack of information on the facility's operational hours; lack of a female midwife; expecting and recent mothers do not know maternal and newborn health care is important; expecting and recent mothers do not think maternal and newborn health care is important; superstitions or other cultural barriers; no family support; and expecting and recent mothers prefer to deliver with and receive pre- and postnatal care from a Traditional Birth Attendant.
56. The second program had a slightly expanded focus on maternal, newborn, and child health: MNCH rather than MNH.
57. See Transparency for Development (2017) and Arkedis et al. (forthcoming).
58. Local councils were created in 2004 in a program sponsored by the World Bank (Acemoglu, Reed, and Robinson 2014).
59. Ghana and Sierra Leone also have structures of traditional authority parallel to other government structures. In the program in Ghana and Sierra Leone, traditional authorities were not formally involved in the program but participants in many communities chose to involve traditional authorities in their efforts.
60. Village Heads are also called village chiefs. In 2017, there were 3,994 Group Village Heads and 264 Traditional Authorities in Malawi (Carlson and Seim 2020). Basurto, Dupas, and Robinson (2020) study a program involving chiefs in two districts in the same region of Malawi as the communities where the program was offered and find that Group Village Heads were responsible for between 2 and 10 villages.
61. See Walsh et al. (2018). For example, traditional authorities played a central role in enforcing a 2012 nationwide ban on delivery with Traditional Birth Attendants in an effort to reduce maternal and child deaths (Butrick et al. 2014).
62. To some degree the change in these interviews before and after the program may reflect interpersonal differences in interpreting the interviewer's question. The increase among participants in Indonesia is less statistically significant once we adjust for each participant's perception of the civic efficacy of people in three hypothetical vignettes trying to improve their school (see Kosack et al. 2019).
63. See Kosack et al. (2019) for analysis and discussion.



A PUBLICATION OF THE

Ash Center for Democratic Governance and Innovation
Harvard Kennedy School
79 John F. Kennedy Street
Cambridge, MA 02138

617-495-0557
ash.harvard.edu



HARVARD Kennedy School
ASH CENTER
for Democratic Governance
and Innovation