Impact of social accountability programs on availability of essential medicines: evidence from Tanzania[⋆,⋆⋆]

Igor Francetic[a,b,c,*], Günther Fink[a,b], Fabrizio Tediosi[a,b]

[a] *Swiss Tropical and Public Health Institute (Swiss TPH), Basel, Switzerland*
[b] *University of Basel, Basel, Switzerland*
[c] *University of Applied Sciences and Arts of Southern Switzerland (SUPSI), Manno, Switzerland*

Abstract

Social accountability tools include different mechanisms to involve communities in management and oversight of public providers. Our research aimed at assessing the impact of a social accountability monitoring program implemented in two 2 of 7 districts in the Tanzanian region of Dodoma on the availability of essential medicines at health facility level. The outcomes considered are the total number of stock-out days for different categories of selected tracer medicines. With a difference-in-differences approach, we estimated an average reduction of up to 80 days of stock-out, depending on the group of drugs considered. The result contributes to the literature on the role of social accountability in health services and availability of essential medicines. This latter issue is highly prioritized in low- and middle-income countries because it affects negatively both patients' experiences and perceived quality of health care, with negative spillover effects on future health seeking behavior and access to care.

Keywords: social accountability, governance, stock-outs, tanzania

1. Introduction

Access to essential medicines is among the top global health priorities, as reflected in the United Nations Millennium Development and Sustainable Development Goals[1]. Despite major efforts in the past decades, making essential medicines available in a consistent and reliable fashion remains a major challenge in many countries. A 2012 UN study on 17 countries showed that between 2007 and 2011 only 51.8% of essential medicines were available in public health facilities at any given point of time (United Nations, 2012).

Stock-outs of essential medicines have been linked to incomplete or lack of immunization (Favin et al., 2012), poor control of non-communicable diseases (Attaei et al., 2017), ineffective antiretroviral therapy (Berheto et al., 2014), , incomplete detection and treatment of malaria (Layer et al., 2014), and increased maternal and child mortality (Githinji et al., 2013). Stock-outs of essential medicines also increase out-of-pocket payments due to purchase from alternative private providers (Mikkelsen-Lopez et al., 2014; Wagenaar

⋆This work was supported by the Swiss Programme for Research on Global Issues for Development [grant nr. 160373], jointly financed by the Swiss National Science Foundation and the Swiss Agency for Development and Cooperation.

⋆⋆The authors are thankful to all r4d research consortium members at NADEL ETH Zurich, Ifakara Health Institute (Tanzania) and University of Ghana for the useful discussions that informed the development of the study. The authors are also grateful to the members of the Health Promotion and System Strengthening (HPSS) project team for the authorization to use their data and the continuous support. We received especially helpful comments from August K. Joachim, Karin Wiedenmayer, Siddhart Srivastava, Fiona Chilunda and Vicky S. Msamba. Finally, we appreciate the useful information and feedback received from Sikika, in the persons of Richard Msittu and Patrick Kinemo.

⋆Corresponding author at: SUPSI DEASS, Via Violino 11, 6928 Manno, Switzerland

Email addresses: igor.francetic@supsi.ch (Igor Francetic), guenther.fink@swisstph.ch (Günther Fink), fabrizio.tediosi@swisstph.ch (Fabrizio Tediosi)

[1] Availability of medicines is included in goals 8E in the MDGs and 3B in the SDGS (see https://sustainabledevelopment.un.org/sdg3).

et al., 2014; Wales et al., 2014), generate dissatisfaction with service and inhibit health seeking behaviour (Ikoh et al., 2009; Kruk et al., 2010; Muhamadi et al., 2010; Tefera et al., 2014). Finally, low availability of essential medicines has been associated with low enrolment in social or community-based health insurance schemes (Fadlallah et al., 2018; Kalolo et al., 2015; Kamuzora and Gilson, 2007; Renggli et al., 2019).

Stock-outs of essential medicines have several and frequently concurrent causes. These include inefficient medicine supply chains, manufacturing shortages, lacking ability to forecast needs and poor inventory management, lack of funding and staffing, weak governance and system fragmentation (Bate et al., 2010; Cameron et al., 2009; Wagenaar et al., 2014; Wales et al., 2014).

A large number of strategies have been put in place in low- and middle-income countries (LMICs) over the last decade to increase the availability of essential medicines at national and sub-national levels. These strategies include supply chain strengthening efforts through improved infrastructure and management systems, specific staff training, information systems and mobile phone text message based systems to monitor stock levels (Barrington et al., 2010; Githinji et al., 2013; Leung et al., 2016) .

Another strategy that has been increasingly used to improve availability of medicines and health system performance more generally are social accountability interventions. Broadly, social accountability aims at improving efficiency of existing systems by allowing patients and citizens to directly provide feedback to health workers. This direct supervision of health workers by their constituencies creates social pressure, which should boost health worker performance in general and improve resource utilization. Conceptually, "social accountability" programs comprise a spectrum of interventions and arrangements set up to hold providers accountable to civil society. In the health sector, accountability and transparency are key elements of health systems governance. Strong accountability mechanisms and high degree of transparency are commonly associated with higher accessibility to services, better quality of care, higher patient satisfaction and provider performance, lower corruption, and ultimately better health outcomes (Greer et al., 2016; Kalolo et al., 2015; Lewis, 2006; Lewis and Pettersson, 2009). In the last decade, a rapidly growing number of social accountability projects applied to health care have been launched in low-income settings (Molina et al., 2017). Successful programs like Uganda's community-based monitoring (Björkman and Svensson, 2009) have fueled interest in these programs as a relatively inexpensive and potentially highly cost-effective way to improve population health. The nature of different social accountability interventions ranges from citizen monitoring and oversight of public sector performance to citizen participation in resource allocation and decision making, public disclosure of data through scorecards, community scorecards and formal community-based health facility committees (Fox, 2015; Joshi, 2017; Joshi and Houtzager, 2012; Molina et al., 2017; Ogbuabor and Onwujekwe, 2018).

Ten years after the rather influential "power to the people" paper (Björkman and Svensson, 2009) the overall evidence on the effectiveness of social accountability programs remains mixed. A more recent study by the same authors still in Uganda suggests that community monitoring may only be effective when monitoring programs are coupled with a scorecards comparing the performance of health providers to others (Björkman Nyqvist et al., 2017). Another study from Uganda shows positive impacts on corruption and mismanagement, as well as on the quality of service provision (Fiala and Premand, 2018); similar results emerged from a qualitative study from India analysing access to maternal health services at the community level (Hamal et al., 2018).

This evidence is contrasted by several studies finding no impact associated with social accountability initiatives (Gaventa and McGee, 2013; Lodenstein et al., 2017a, 2018). The main reasons mentioned for the lack of success are the mistrust towards the authorities for monitoring programs embedded in formal government structures or the involvement of citizens in monitoring policies that do not contribute to shape. For example, institutionalized health facility committees in Tanzania have been found to have no relevant effect on governance or service provision (Frumence et al., 2014; Kessy, 2014; Kigume and Maluka, 2018; Macha

et al., 2011). Health facility committees showed no association with health financing and service provision in Nigeria (Ogbuabor and Onwujekwe, 2018). Feruglio and Nisbett (2018) also show that community-level accountability mechanisms in India do not involve communities and are often captured by frontline workers included in these committees. Other factors that may limit the effectiveness of social accountability initiatives are lack of solid legal frameworks, lacking prosecution of detected misconduct (McNeil and Male, 2010) as well as the lack of consideration of both actors (providers and patients) of the accountability relationship (Gaventa and McGee, 2013).

In this study, we assess the impact of a social accountability monitoring (SAM) program in Tanzania. The program was implemented by a local NGO named Sikika in two out of seven districts of Tanzania's Dodoma region. These two districts where chosen based on historical links Sikika had established with local grassroots level civil society organizations (CSOs). The treatment districts were similar to the five other districts of Dodoma region in terms of demographic, socio-economic development, and health systems characteristics at baseline. The main objectives of the SAM program were to increase the availability of medicines at health facilities and to improve health facilities infrastructure maintenance, resource allocation and financial management at the district level. A last area of interest for Sikika was the performance assessment of existing oversight and accountability bodies, with a specific focus on their ability to hold officials accountable and promote improvements in health service delivery. (Sikika, 2013) The SAM program was implemented starting from 2012, in parallel with a large health system strengthening effort (the Health Promotion and System Strengthening Project, HPSS) run in the entire Dodoma region, through which detailed information on facility quality and functioning was collected.

Using a difference-in-differences approach to identify the additional improvements in the SAM districts over the periods 2011 and 2017 we find that the SAM intervention was relatively successful in reducing the number of stock-out days. Considering the pool of all drugs included in the analysis, on average the SAM intervention reduced the number of drug-stockout days in the 3 months period preceding the survey by 80 days (out of 1170 drug/days), which corresponds to about 0.54 standard deviations of stockout days observed at baseline. These reductions seem to be most pronounced among antibacterial drugs. The same positive impacts were not seen for infrastructure maintenance, which was also targeted by the same program.

In the last part of the paper, we explore the potential mechanisms that explain the incremental effect of the SAM project. Overall, it appears that the SAM intervention did not only strengthen the links between health facility staff and local communities through health facility committees, but also strengthened the link between district authorities and health facility representatives. The results presented in this study contribute to a growing body of evidence suggesting that increased transparency (or greater access to public information) alone may not necessarily improve outcomes or may only improve outcomes in settings where marginal changes in provider efforts can affect observable outcomes. (Fiala and Premand, 2018; Fox, 2007, 2015; Joshi, 2017; Joshi and Houtzager, 2012; O'Meally, 2013) The main enabling factor highlighted in the literature that is supported by our analysis is context specificity (Danhoundo et al., 2018; Gaventa and McGee, 2013; Martin Hilber et al., 2016). Another crucial aspect pointed by the literature is the existence of solid loops for feedback between community and government authorities (Fox, 2015; Ringold et al., 2012). To this extent, our interpretation is that the multiple successful rounds of discussion between the SAM teams and district authorities as well as the feedback to communities brought by Sikika may be important pathways to impact. Further facilitating factors highlighted in the literature and potentially relevant for the Sikika example are leveraging existing partnership, building coalitions, having clearly defined roles for citizens and civil servants involved (Danhoundo et al., 2018), addressing appropriately both demand and supply-side, acting strategically and politically exploiting "windows of opportunity" (Dewachter et al., 2018; O'Meally, 2013).

Our paper contributes to the above literature by providing medium run evidence on the impact of a well-defined social accountability intervention - managed outside of formal government processes - that includes elements of information and feedback to community involving demand and supply-side stakeholders.

The remainder of the paper is structured as follows: section 2 describes the general setup and local health systems, section 3 discusses data and methods. We present our main results in section 4 and our conclusions in section 5.

2. Setup

2.1. Study setting

Stock-outs of essential medicines have been a central health system problem in Tanzania for many years. A baseline survey conducted in 2002 across 4 regions showed that 75% of surveyed health facilities reported 28 or more days of stock-out across a set of essential medicines (Ministry of Health, 2002). Another survey conducted at national scale in 2006 found that two thirds of government managed health facilities could not abide to the treatment protocol for malaria as a result of shortage of medicines (National Bureau of Statistics and Macro International Inc., 2007). The same survey also showed that almost 50 percent of public health facilities in Tanzania did not have all first-line treatment drugs for TB in stock. Likewise, 48 percent of public health facilities in the country reported stock-outs in normally stocked antiretroviral drugs (National Bureau of Statistics and Macro International Inc., 2007). More recent surveys conducted in 2012 and 2014/2015 essentially confirmed the earlier figures, revealing full availability in less than half public health facilities (MoHSW and Ifakara Health Institute, 2013; MoHSW et al., 2016). Specific data for the Tanzanian Region of Dodoma revealed that in 2011 the availability of 24 tracer medicines in public health facilities was as low as 53% (Wiedenmayer et al., 2019).

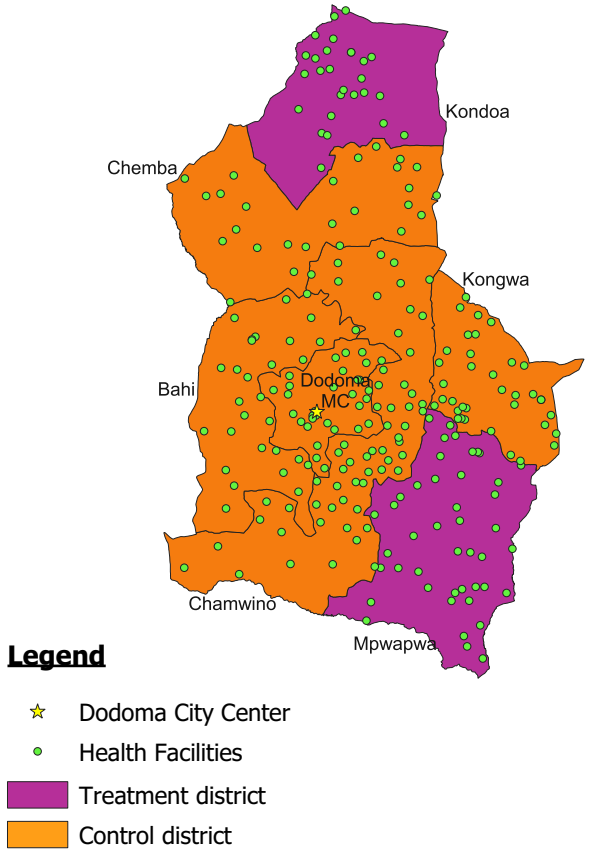
Over the last decade, the Government of Tanzania has adopted a series of specific reforms to increase the availability of essential medicines (Mujinja and Kida, 2014). These include: (1) a shift from a "push" delivery system of predetermined kits to a "pull" system based on local needs through an Integrated Logistic System (ILS) managed by the Ministry of Health; (2) greater coordination across institutions through a sector-wide approach with technical working groups involving stakeholders; (3) increased monitoring and data collection; (4) improved budgeting and procurement processes; (5) improved prescription practices and rational use of essential medicines; (6) decentralization of management to local government authorities for increased ownership and better prioritization of local needs; (7) introduction of complementary supply (e.g. Accredited Drug Dispensing Outlets, Prime Vendor system) from the private sector to fill the capacity gap of the public system (Musau et al., 2011; Wales et al., 2014). Despite these efforts, substantial gaps remain, with major stock-out periods at most facilities surveyed in this study at baseline.

This study was conducted in Dodoma, one of the 31 regions of Tanzania. The Dodoma region includes and shares the name with the current capital city of Tanzania, Dodoma City. The area surrounding Dodoma City represents the only predominantly urban district of the seven composing the Dodoma region, the others being Bahi, Chamwino, Chemba, Kondoa, Kongwa and Mpwapwa. Figure 1 shows a map of the region with districts borders. As of 2012, Dodoma had a population of about 2.1 million people spread across 41'311 square km; 16% of the population lived in urban areas. Life expectancy at birth was 49 and 51 years for females and males, respectively, substantially below the national averages of 64 (females) and 61 (males). Under five mortality rate was 76.4 per 1000 live births in 2012, slightly above the national average of 66 (National Bureau of Statistics and Office of Chief Government Statistician, 2013).

Starting in 2012, a Health Promotion and System Strengthening (HPSS) program was implemented in the entire region (University Consultancy Bureau, 2018). Supported by the Swiss Agency for Development and Cooperation the HPSS program includes five main components: Health Promotion, Health Financing, Medicine Management, Health Technology and Maintenance and crosscutting issues such as gender, social inclusion and HIV/AIDS[2]. Acknowledging the systemic roots of the problem (Bigdeli et al., 2013; De Savigny

[2]See http://www.hpss.or.tz and https://www.jaziaprimevendor.or.tz/ for further details.

Figure 1: Districts in Dodoma region and location of surveyed health facilities.



Note: source for GIS layers with district boundaries are DHS and NBS (2017). GIS position for surveyed health facilities was recorded by HPSS at the time of the survey (2011 and 2017).

and Adam, 2009), the project's Medicine Management component has specific sub-components addressing health workers training, accountability, rational use of medicines and notably medicines supply-chain (Wiedenmayer et al., 2019). This latter component is explicitly meant to address stock-outs, inefficiencies and shortages in medicines deliveries attributable to the centralized government central Medical Store Department (MSD) and identified as major cause of stock-outs.

The medical supply chain tool - piloted from September 2014 to July 2018 in all districts in the Dodoma Region - is a complementary supply of medicines through an innovative public-private partnership initiative known as "Jazia Prime Vendor" (PVS) system (Stoermer, 2017). The PVS allows district authorities to purchase medicines from a single private provider when the centralized MSD experiences shortages. The single private supplier – the Prime Vendor – is tendered and contracted for the whole region based on quality attributes. District authorities are then responsible for distribution to health facilities of supplies purchased through the PVS. Before the PVS implementation in Dodoma, in case of MSD shortages each district authority could purchase complementary medicine supplies from different private providers without specific public tender procedures. The resulting system was bureaucratic, expensive and more exposed to corruption. (Wiedenmayer et al., 2019)

Besides strengthening the capacity of existing health facility committees, HPSS does not include specific social accountability elements. The HPSS project – with all its components - was rolled out simultaneously to all districts across the region since the year 2012 (University Consultancy Bureau, 2018).

2.2. The intervention: Social Accountability Monitoring

The social accountability intervention evaluated in our study was implemented by a local NGO called Sikika. Sikika's stated mission is "to enhance health and public finance systems through Social Accountability and advocacy at all government levels"[3]. With funding from several development partners, Sikika launched a Social Accountability Monitoring (SAM) program targeting health service delivery in 2012. The program was implemented in two out of the seven districts in Dodoma: Kondoa and Mpwapwa. These two districts were chosen based on existing strong partnerships with grassroots organizations. Geographically, Kondoa and Mpwapwa lie at the Northern and Southern ends of the region, with all other (control) districts located in between these two areas.

In Table 1 we show baseline average differences between control and treatment groups across a series of relevant population, health system and health facilities characteristics. The sensible difference in health insurance coverage in the treatment group can be explained by the specific policy – in Kondoa district - of automatically enrolling all people covered by a poverty alleviation scheme (Tanzania Social Action Fund, TASAF) in the community health insurance scheme (Sikika, 2016). Therefore, it should not reflect differences in dimension relevant for our study. The data also show higher frequency and duration of stock-outs in the treatment group, compared to the control group. However, accounting for error clustering at district level, the reported differences are not statistically significant. Overall, we can thus claim that groups do not differ systematically in terms of their economic development or health systems features.

Sikika's SAM project consists of an articulated accountability process that – in line with the recommendations found in the literature - involves several different steps: formation and training of community based supervision teams, district stakeholders meetings, field visits for data collection, analysis, reporting to district authorities, feedback to stakeholders and continuous monitoring.

Before starting the implementation of the SAM process in a new district, Sikika meets with Local Government Authorities and other stakeholders to introduce the SAM process and principles of the project. After this preliminary stage, community meetings are held to select democratically citizens that will join the

[3]See http://www.sikika.or.tz for further details.

Table 1: Differences between Control and Treatment group at baseline

Variable	Control group Obs.	Control group Avg./Perc.	Treatment group Obs.	Treatment group Avg./Perc.	Difference (C - T)
Health insurance coverage (perc.)	243	14.40	94	24.47	-10.07
Children sleeping under bednet (perc.)	212	76.41	65	70.77	5.64
Delivery at health facility (perc.)	217	37.78	91	36.26	1.52
Children birth weight (g)	94	3215.43	37	3213.51	1.91
Anemic children (perc.)	134	51.49	45	53.33	-1.84
Children received BCG vaccine (perc.)	206	94.66	89	95.50	-0.84
Children received first DPT vaccine (perc.)	206	93.20	89	94.38	-1.18
Children received Polio vaccine (perc.)	206	77.67	89	78.65	-0.98
Children received Measles vaccine (perc.)	205	74.17	88	69.32	3.85
Active health facility committee (perc.)	146	91.78	59	88.13	3.64
Facilities located in urban area (perc.)	146	6.85	59	0.00	6.85
Share of pop. in 1st wealth index quintile (perc.)	146	35.6	59	33.33	2.3
Share of pop. in 5th wealth index quintile (perc.)	146	5.73	59	0.8	4.93
Average yearly rainfall (mm per year)	146	548.68	59	624.94	-76.26
Malaria prevalence (parasite rate)	146	6.21	59	4.70	1.51
Stock-out days: antibiotics	146	62.66	59	89.90	-27.24
Stock-out days: antimalarials	146	51.05	59	48.10	2.95
Stock-out days: other drugs/vaccines	146	120.11	59	142.86	-22.75
Stock-out days: reproductive health	146	44.53	59	57.44	-12.91
Stock-out days: all drugs	146	192.69	59	239.39	-46.70
Index for maintainance of storage area	146	7.64	59	8.01	-0.37
Index for maintainance of dispensing area	146	5.56	59	6.56	-1.00

Notes: (1) Urban area, malaria prevalence, average yearly rainfall and wealth quintile attributed to health facilities matching characteristics of the closest survey cluster. (2) Distance from MSD computed based on shortest road distance. (3) Detailed sources are listed in appendix A3. (4) All differences are not statistically significant under test for differences in means or proportions with errors clustered by district.

district SAM team, composed of 15-20 members overall. District SAM teams gather citizens, representatives from district authorities (Council Health Management Team, Council Management Team), Health Facility Governing Committees (HFGCs), religious groups, local NGOs or other grassroots level CSOs. Each SAM team is responsible for the implementation of the SAM process, covering all government managed health facility within the district boundaries (in our case 20-50 depending on the district). Once the formation of SAM teams is complete, members follow a two weeks training on various topics, including principles of SAM, human resources for health, professional integrity, planning and resource allocation, expenditure management, performance management and oversight bodies. The full SAM process encompasses the following steps for each round, taking about one and a half years each:

1. Issues identification and analysis of relevant documents to prepare site visits at Health Facilities. Documents analysed include district strategic plans, comprehensive council health plans that include budget allocations (for infrastructure, medicines, human resources, etc.), implementation plans in the domain of health and healthcare as well as internal audit reports;
2. Based on the issues identified, SAM teams visit a sample of health facilities across the district to assess the situation, in terms of infrastructure investments and maintenance, disbursement and utilization of funds, availability of medicines, human resources, management capacity and functioning of HFGC;
3. After site visits, analysis and reporting to Council Health Management Team (CHMT) with preliminary results and questions (feedback due within 1 week);
4. Internal SAM team meetings to discuss questions and answers emerging from the reporting. The SAM team discusses issues and evidence collected across the district, looking for an agreement on

improvements needed;

5. Meeting with district stakeholders to discuss results of SAM;
6. Community meetings to discuss results of SAM, presenting issues identified and responses received from district authorities to citizens, community representatives and health workers, looking for a consensus about the issues;
7. Definition of a monitoring strategy for the next years. Citizens and monitors are engaged, taking into account their respective capacity. CHMT updated on the progress;
8. Meetings between monitors and SAM teams on a monthly basis. Regular updates to the community about progress;
9. Participation in planning and budget allocation with district authorities.

The outcomes targeted by the program were:

- Reduction of stock-outs in essential medicine
- Improvement of infrastructure maintenance
- Improvement of the allocation, disbursement and utilization of funds received from government basket fund, Community Health Fund (CHF) and National Health Insurance Fund (NHIF)
- Improvement of existing local governance and accountability arrangements (eg. Health Facility Governing Committees, or HFGCs)

The intermediary effects expected from the SAM project were increased community sensitization, engagement and empowerment, improved accountability of health workers and district representatives as well as increased transparency, i.e. effective disclosure and access to public district budget documents and plans for SAM teams and more generally all interested citizens. Sikika also developed and implemented a multi-channel media strategy aimed at facilitating the achievement of the above objectives. (Sikika, 2013)

The SAM intervention presented above was launched in 2012 and thus implemented in parallel with the HPSS project. Given this, our study identifies the incremental effect of SAM in the context of larger health system reform efforts. Appendix A1 describes the detailed timing of the implementation of HPSS and SAM programs.

3. Data and methods

3.1. Data

The main data sources for our analysis are two health facility surveys conducted in 2011 and 2017 within the region-wide HPSS project. HPSS conducted households and health facilities surveys at baseline in 2011 and endline in 2017. Both health facilities surveys covered only government-managed health facilities; private and faith-based facilities were neither included in the HPSS project nor addressed by the SAM program. The facility questionnaires explored different aspects of health service provision at health facility level related to the components addressed by the project. As of 2012, the Dodoma region had 267 government-managed health facilities. The number rose to about 286 in 2017 (Ministry of Health and Social Welfare [MoHSW], 2011; Ministry of Health, Community Development, Gender, Elderly and Children [MoHCDGEC], 2019). The 2011 baseline HPSS survey included all government-managed health facilities in the region. For the 2017 endline survey, HPSS sampled randomly about half of the health facilities in the region, stratifying by district and health facility type (University Consultancy Bureau, 2018). The sample resulting from preliminary data cleaning procedures includes 226 health facilities, 205 of which observed at baseline. Excluding health facilities observed only at endline, the randomly selected group of facilities observed for follow-up represent 44% of health facilities (N=91) in our data. The sample is balanced over time across types of facilities, groups (control and treatment), and districts (see Table 2 and Appendix A2).

Table 2: Distribution of health facilities and follow-up rates in control and treatment groups

Baseline (Pre)				
	Dispensary (*count and %*)	Health centre (*count and %*)	District hospital (*count and %*)	Total (*count and %*)
Control	124	19	3	146
	84.93	13.01	2.05	100.00
Treatment	54	3	2	59
	91.53	5.08	3.39	100.00
Total	178	22	5	205
	86.83	10.73	2.44	100.00

Endline (Post)				
	Dispensary (*count and %*)	Health centre (*count and %*)	District hospital (*count and %*)	Total (*count and %*)
Control	72	8	3	83
	86.75	9.64	3.61	100.00
Treatment	24	3	2	29
	82.76	10.34	6.90	100.00
Total	96	11	5	112
	85.71	9.82	4.46	100.00

Notes: (1) Tests confirm no significant difference in proportions accross groups and over time. (2) Control group: 65 of 146 facilities at baseline randomly sampled for follow-up (44.52 %). (3) Treatment group: 26 of 59 facilities at baseline randomly sampled for follow-up (44.06%). (4) Crude follow-up rates (Control = 56.85%, Treatment = 49.15%) include facilities observed only at endline for both groups. (5) Tests confirm that crude and net follow-up rates arestatistically equivalent across groups.

We complemented the HPSS survey data with multiple additional sources. Annual Health Statistics developed by the Tanzanian Ministry of Health and Social Welfare (MoHSW) and National Bureau of Statistics (NBS) were used to compute district-level numbers on population, density of health facilities, outpatient visits as well as the prevalence of tuberculosis and malaria. To control for household level trends in wealth and morbidity, we also construct district and cluster level averages using Demographic and Health Surveys (DHS) from 2010, 2011, 2015 and 2017. We include a detailed list of data sources in Appendix A3.

The primary outcome measure of interest were medicines stock-outs days. The HPSS survey collected detailed data on 22 essential medicines, which we use as our primary outcome (University Consultancy Bureau, 2018). Stock-out days represents the sum of days out of stock (duration) over the selected medicines for the 3 months prior to the survey. Reducing days of stock-out – besides reducing their frequency – has major implications for population health. In fact, the effectiveness of pharmaceutical treatment of several health conditions depends crucially on the timing of the treatment itself. It follows that the days of reduction in stock-out days can be broadly interpreted as increase in effective treatment days. Unfortunately, data limitations restricted our analysis to a subset of 13 items included in the HPSS tracer medicines list. Table 3 lists the composition of drug groups included in our analysis.

Given that the SAM program targeted four outcomes, we also analysed program impact on infrastructure maintenance of medicines storage and dispensing areas. Our indexes of maintenance of storage and dispensing areas result from 16 different questions listed in Appendix A4. Even though some data on funding flows were collected, the data was too incomplete to allow for difference-in-difference analysis.

We included several contextual factors – at health facility, cluster and district level - that may affect availability of drugs in our model. At health facility level, we included the type of health facility, the existence of a functioning HFGC and the distance to the city centre of Dodoma as proxy for the distance

Table 3: Overview of drugs considered in the analysis and composition of sub-groups

Drug type	Main indication	Sub-groups			
		Antibiotics	Antimalarials	Other drugs and vaccines	Reproductive health
Artemether/lumefantrine	Antimalarial	-	Yes	Yes	-
Quinine	Antimalarial	-	Yes	-	-
Amoxicillin 250mg caps	Antibiotic	Yes	-	Yes	-
Amoxicillin syrup	Antibiotic	Yes	-	Yes	-
Benzyl Penicilling 5MU inj	Antibiotic	Yes	-	Yes	-
Metronidazole tabs	Antibiotic	Yes	-	-	-
ORS sachet	Diarrhea	-	-	Yes	-
Paracetamol 500mg tabs	Fever and pain	-	-	Yes	-
Medroxyprogesterone inj	Contraceptive	-	-	-	Yes
Oxytocin	Contraceptive	-	-	-	Yes
Ferrous salt and folic acid	Anemia	-	-	-	Yes
Condoms	Contraceptive	-	-	-	Yes
Vaccines	Immunization	-	-	Yes	-
Maximum days of stock-out	1170	360	180	630	360
Number of items	13	4	2	7	4

from the central zonal MSD store (Wagenaar et al., 2014). This latter variable does not vary over time. At cluster level, we included a dummy variable that indicates urban areas (time invariant), the share of population in 1st and 5th wealth index quintiles and the average yearly rainfall and malaria prevalence. In this context, cluster average rainfall is a measure of ease of access to health facilities during rainy season (Adhvaryu and Nyshadham, 2015; Moïsi et al., 2010). All cluster variables are attributed to health facilities according to the geodesic distance between cluster point and health facility coordinates. At district level, we controlled for district population, health facility density, prevalence of tuberculosis and volume of outpatient visits.

Table 4 shows selected descriptive statistics for our outcomes and control variables at baseline (2011). The majority of health facilities in our sample were located in rural areas (about 95%) and had an active health facility committee (91%). The distribution of health facilities across districts within the region was relatively similar, with an average of about 17 health facilities per 100'000 population (range 15.27 to 18.04). District population ranged from 210'682 to 410'956, with an average size of 294'395 people in 2011. The outcomes of interest are listed in the second part of the table. Notably, for two categories (antimalarials and drugs prescribed in reproductive health), at least one facility reported being out of stock for the full three months prior to the survey dates. The average number of stock-out days ranges between 13 and 27 percent of the maximum potential – described in Table 3 - across all classes of medicines considered.

3.2. Empirical approach

Our study aims at assessing the impact of SAM on health facility performance, measured as duration of stock-out in 13 essential medicines. Our main difference-in-differences (DID) model explores differences between treatment and control areas over time. Our treatment group includes the districts of Kondoa and Mpwapwa, where the Sikika program was rolled out. The control group includes the other five districts from the same region. As Figure 1 shows, these five districts - Bahi, Chemba, Chamwino, Dodoma urban and Kongwa – are spatially located in between the two treated areas.

Table 4: Descriptive statistics for all variables at baseline (2011)

	N	Avg./Perc.	Std. Dev.	Min	Max
Active health facility committee (perc.)	205	90.73		0.00	100.00
Urban area (perc.)	205	4.88		0.00	100.00
Population in 1st wealth index quintile (perc.)	205	34.97		0.00	78.16
Population in 5st wealth index quintile (perc.)	205	4.31		0.00	73.28
Distance from MSD store (meters)	205	103289.41	55966.84	1868.62	237086.69
Average rainfall around facility (mm per year)	205	570.63	109.92	424.92	814.00
Malaria prevalence around facility (parasite rate)	205	0.06	0.02	0.03	0.10
District population (number)	205	294395.63	58204.31	210682.00	410956.00
District health facility density (per 100'000)	205	16.83	0.89	15.27	18.04
District TB notification rate (per 100'000)	205	94.25	47.40	56.00	235.30
Number of OPD visits (number)	205	12381.35	4342.02	6173.64	18389.16
Antibiotics (stock-out days)	205	70.50	71.96	0.00	320.00
Antimalarials (stock-out days)	205	50.20	48.49	0.00	180.00
Other drugs and vaccines (stock-out days)	205	126.66	92.21	0.00	453.00
Drugs for reproductive health (stock-out days)	205	48.25	63.77	0.00	360.00
All drugs (stock-out days)	205	206.13	147.92	0.00	723.00
Index for maintainance of storage area	205	7.75	2.62	0.00	14.00
Index for maintainance of dispensing area	205	5.85	2.42	0.00	12.00

Notes: (1) Urban area, malaria prevalence, average yearly rainfall and wealth quintile attributed to health facilities matching characteristics of closest DHS cluster. (2) Distance from MSD computed based on shortest road distance. (3) Indexes forinfrastructure maintainance of medicines storage and dispensing result from a checklist of 16 items included in the HF survey. (4) Detailed sources for all data are listed in appendix A3.

The estimated difference-in-differences model is given by:

$$Y_{i,d,t} = \beta_0 + \beta_1 Post_t + \beta_2 Treatment_d + \delta Post_t \times Treatment_d + \phi_{i,t} X_{i,t} + \gamma \Omega_{d,t} + \epsilon_{i,d,t} \quad (1)$$

where $Y_{i,d,t}$ is the outcome of interest for health facility i in district $d \in D$ for period $t \in 0 = 2011, 1 = 2017$). $Treatment_d$ is an indicator variable that equals 1 for health facilities in the treatment districts and zero for those included in the control group. $Post_t$ is a binary indicator for the time period (either 0 for the pre-treatment 2011 baseline or 1 for the post-treatment 2017 endline). $X_{i,t}$ is a matrix of health facility (individual and cluster) control variables, $\Omega_{d,t}$ is a matrix of district level control variables, ϕ and γ are the vectors of coefficients for the control variables. Finally, $\epsilon_{i,d,t}$ is the idiosyncratic error term. Our coefficient of interest - estimating the effect of the SAM program - is represented by δ.

As long as the common trends assumption is satisfied – which we will explore in the section devoted to robustness checks - the coefficients in model 1 should be consistently estimated with ordinary least squares (OLS). However, the small number of clusters poses a challenge in obtaining unbiased standard errors for the estimated coefficients (Wooldridge, 2010). To deal with the small number of clusters, we first perform significance tests using a t distribution with $G-1$ degrees of freedom (where G is the number of clusters, in our case $G = 7$) following Donald and Lang (2007). We also explore inference tests based on wild clustered bootstrap (Cameron et al., 2008) with a 6-point weight distribution (Webb, 2013; Rokicki et al., 2018). Finally, we explore sub-cluster wild bootstrapping as recommended by MacKinnon and Webb (2018).

4. Results

4.1. Analysis results

The results of our preferred model in 1 for stock-out days are presented in table 5; standard errors in our main model are computed using cluster robust standard errors (CRSEs). The analysis was conducted

using the statistical package Stata, version 14. The estimated impact of the Sikika SAM programme is a significant reduction in stock-out days of about 59 days (out of a potential maximum of 360 and a baseline average of 70.50) for antibiotics and about 80 days (out of 1170, baseline average 206.13) for all drugs. The result for antibiotics is remarkable if we consider that SAM decreases days of stock-out by about 84% of the baseline average (0.83 standard deviations) and given a baseline median value of 60.

Our regression results also show a reduction in stock-out days associated to the SAM program for drugs prescribed for reproductive health (about 23 days), other drugs and vaccines (about 45 days) and an increase for antimalarial medicines (about 3 days), although the effects are not statistically significant at the 95% confidence level.

Conditionally on all control variables, we find consistent and significant reduction in stock outs in the post time period for all categories except for medicines used in reproductive health; these positive trends likely reflect at least to some extent the efforts of the region-wide HPSS project.

Table 6 includes t-statistics and p-values under our three alternative approaches that address the excessive rejection rates of the null hypothesis associated to CRSEs (t distribution with 6 degrees of freedom, wild clustered bootstrap and wild bootstrap at sub-cluster level). Following Roodman et al. (2018) – due to the strong assumption of asymptotic normality of coefficients required - we do not provide the standard errors implied by the t statistics obtained under the wild clustered bootstrap approach. Accordingly, the last column in table 6 shows the p-value inflation factor compared to standard CRSEs.

Under CRSEs, the policy effect is statistically significant for antibiotics and all drugs pooled. The alternative test on the t statistic computed from standard CRSEs deflates p-values across all outcomes by a factor of 0.5. This results in relaxed inference criteria compared to the standard CRSE approach; using a cut-off value of 5%, the policy effect is significant under alternative t test approach for all drug groups, except antimalarial drugs.

The p-values under wild cluster bootstrap are sensibly larger compared to standard CRSE[4]. Under wild cluster bootstrap, our SAM policy is significant at the 5% confidence level only for antibiotics.

The third and last alternative approach that we applied to correct inference accounting for the effect of clustering with a small number of clusters is the wild sub-cluster bootstrap procedure. Compared to the wild cluster bootstrap, the sub-cluster approach is generally more generous. Using the usual cut-off value of 5%, the program impact is significant for stock-out days of antibiotics, generic drugs/vaccines and all drugs pooled. Interestingly, sub-cluster bootstrap deflates p-values associated to the policy coefficient estimated for other drugs and vaccines (0.69 times the value observed for CRSE).

The comparative inference exercise in table 6 reveals that the simple solution to account for small number of clusters suggested by Donald and Lang (2007) is likely not conservative enough in a setting with few clusters and only two time periods. On the other hand, whilst the penalization on p-values under wild cluster bootstrap may be too strong and result in very strict inference criteria, the sub-cluster bootstrap produces results comparable to CRSE and thus suboptimal.

We applied the same comparative method with three alternative inference approaches on the coefficient representing the time trend in stock-out days (Post variable), that broadly captures the effect of HPSS activities over time. The effect is negative and significant across all inference approaches for stock-out days

[4]Notably, for our measure of stock-out days for antibiotics, the probability of false positive resulting from the bootstrapped t distribution obtained through wild cluster bootstrapping is almost 63 times larger compared to the corresponding CRSE p-value.

Table 5: Diff-in-Diff specification for groups of drugs and pooled variable over period 2011-2017

	Number of stock-out days for				
	(1) Antibiotics	(2) Antimalarials	(3) Other drugs and vaccines	(4) Reproductive health	(5) All drugs
Treatment group	59.67**	15.34	45.53	-21.77*	49.34
	(13.08)	(12.15)	(19.33)	(7.18)	(23.54)
Post	-34.92*	-22.01*	-79.27***	-18.32	-113.21***
	(9.84)	(8.35)	(10.63)	(8.45)	(13.55)
Post × Treatment group	-59.21***	3.04	-45.18	-23.63	-79.95*
	(9.58)	(6.79)	(18.49)	(11.17)	(25.63)
Health centre	10.37	5.93	10.66	15.69	36.47
	(9.28)	(5.51)	(9.54)	(12.83)	(22.76)
District hospital	-50.68*	-14.97	-90.66**	11.80	-79.89
	(13.84)	(11.95)	(21.99)	(32.96)	(52.89)
Active HF committee	21.62	5.88	24.07	-11.97	18.52
	(9.11)	(9.36)	(15.45)	(14.88)	(27.15)
Distance from MSD store	-0.00*	-0.00	-0.00*	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Urban area	7.34	-7.63	15.09	-28.50	-40.30
	(11.44)	(18.41)	(26.96)	(13.46)	(36.45)
Share of pop. in 1st wealth quintile	-37.53	-18.94*	-51.37	-12.44	-75.93
	(29.84)	(7.19)	(32.11)	(17.15)	(50.80)
Share of pop. in 5st wealth quintile	-71.12*	-13.52	-77.43**	9.17	-80.29*
	(20.27)	(14.98)	(15.41)	(13.36)	(26.96)
Average yearly rainfall	-0.05	-0.06	-0.10	0.01	-0.06
	(0.07)	(0.04)	(0.06)	(0.05)	(0.13)
Malaria prevalence	-55.00	-143.12	-336.98	-199.41	-197.71
	(267.58)	(315.13)	(617.28)	(265.55)	(946.86)
District population	0.00**	-0.00**	0.00**	0.00*	0.00**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Health facility density	-7.78	-5.05	-4.10	3.76	-5.70
	(4.56)	(2.84)	(8.53)	(2.89)	(8.98)
TB notification rate	-0.63*	0.19	-0.53*	-0.43*	-1.08**
	(0.19)	(0.08)	(0.20)	(0.12)	(0.29)
Nr. of OPD visits	-0.01**	0.00***	-0.01*	-0.01**	-0.01**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Adjusted R^2	0.16	0.15	0.17	0.02	0.09
AIC	3515.05	3273.81	3681.99	3444.81	3977.85
BIC	3537.61	3296.36	3704.54	3467.36	4000.41
Observations	317	317	317	317	317
Baseline average	70.50	50.20	126.66	48.25	206.13
Baseline standard deviation	71.96	48.49	92.21	63.77	147.92

Cluster robust standard errors in parentheses; * < 0.05, ** < 0.01, *** < 0.001

Notes: (1) The estimation is performed using ordinary least squares. (2) Stock-out days are computed over the 3 months prior to the survey dates (September 2011 for baseline and May 2017 for endline).

Table 6: Results for tests on policy coefficient with alternative inference strategies

Variable	Test	DoF	t Statistic	P-value	P-value Inflation factor
Antibiotics	Alternative t test	6	6.18	0.00 ***	0.50
Antibiotics	Wild cluster	6	6.18	0.05 *	59.31
Antibiotics	Subcluster	225	3.64	0.00 ***	1.21
Antimalarials	Alternative t test	6	0.45	0.33	0.5
Antimalarials	Wild cluster	6	0.44	0.72	1.08
Antimalarials	Subcluster	225	0.28	0.76	1.14
Other drugs/vaccines	Alternative t test	6	2.44	0.02 *	0.50
Other drugs/vaccines	Wild cluster	6	2.44	0.16	3.15
Other drugs/vaccines	Subcluster	225	2.11	0.03 *	0.69
Reproductive health	Alt t test	6	2.11	0.04 *	0.50
Reproductive health	Wild cluster	6	2.11	0.22	2.80
Reproductive health	Subcluster	225	1.61	0.11	1.36
All drugs	Alternative t test	6	3.12	0.01 *	0.50
All drugs	Wild cluster	6	3.12	0.12	6.07
All drugs	Subcluster	225	2.34	0.02 *	1.16

Cluster robust standard errors in parentheses; * < 0.05, ** < 0.01, *** < 0.001

Notes: (1) lines represent significance tests for DID coefficients (Post × Treatment) estimated in Table 5, for the different drug categories considered in our analysis (antibiotics, antimalarials, other drugs and vaccines, drugs for reproductive health, all drugs pooled); (2) The three approaches to inference test are: t distribution with $G-1$ degrees of freedom (where G is the number of clusters), wild clustered bootstrap and sub-cluster wild bootstrap. (3) The empirical approach is described in detail in section 3.

in antibiotics, generic drugs/vaccines and the variable pooling all categories of medicines. A table with all specific results and p-values is provided in Appendix A5. This result favors the hypothesis that the system-wide efforts put in place by HPSS, including the specific Medicine Management components of the project, resulted in a reduction of stock-out days over our study period.

Finally, the regression results of the analysis on our secondary set of outcome variables, related to infrastructure maintenance, are included in table 7. We ran the same model specifications as for drugs but our analysis did not detect any impact on these outcomes. This is unfortunate as it suggests that the positive impact of the Sikika SAM program is limited to the domain of medicines availability.

4.2. Robustness checks

4.2.1. Testing the common trends assumption

Our empirical approach rests on the key assumption of parallel trends in treatment and control groups in absence of the treatment (Angrist and Pischke, 2009; Dimick and Ryan, 2014). Below we present two checks conducted to assess the plausibility of the parallel trend assumption. In detail, we used two different indicators to show that the treatment and control districts do not display systematic differences, for example related to differential governmental efforts.

First, we assessed the share of children under five sleeping under insecticide-treated bed nets (ITNs). Under the National Treated Nets Programme (NATNETS), pregnant mothers in Tanzania can access discounted ITNs through vouchers distributed during the first visit to a Reproductive and Child Health clinic (Eze et al., 2014). The voucher scheme for ITNs targets pregnant mothers and their infants under five, the groups at higher risk from malaria. The scheme requires health facilities to obtain vouchers from district authorities, much like they obtain drug deliveries from the zonal MSD. Data from DHS surveys from 2007 to 2011 (left box of figure 2) show remarkably similar patterns in utilization of ITNs for children aged five

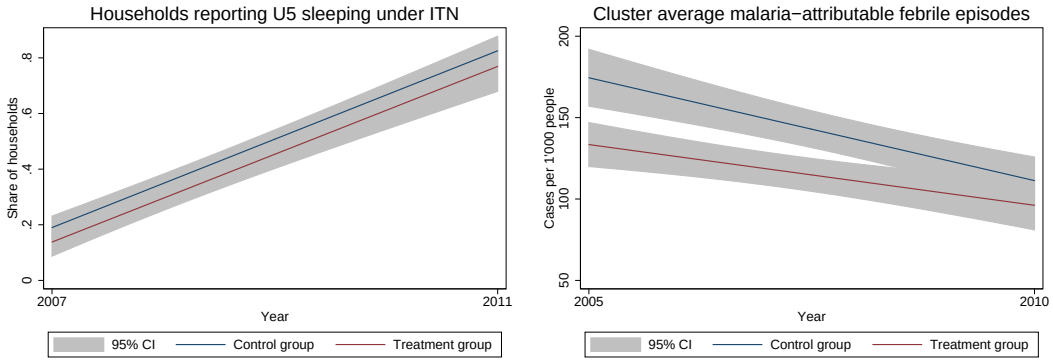
Table 7: Diff-in-Diff specification on infrastructure maintenance variables over period 2011-2017

	(1) Maintenance of storage area	(2) Maintenance of dispensing area
Treatment	0.25	0.94
	(1.02)	(1.29)
Post	4.13*	7.68**
	(1.34)	(1.49)
Post × Treatment	-1.18	-0.96
	(0.92)	(1.16)
Health centre	1.16	0.93
	(0.87)	(0.46)
District hospital	1.40**	2.51*
	(0.29)	(0.85)
Active HF committee	1.21	0.58
	(1.04)	(0.27)
Distance from MSD store	-0.00	-0.00
	(0.00)	(0.00)
Urban area	1.02	0.52
	(0.60)	(0.53)
Share of pop. in 1st wealth quintile	0.49	2.65*
	(0.76)	(0.82)
Share of pop. in 5st wealth quintile	0.43	0.94
	(1.01)	(1.02)
Average yearly rainfall	0.00	0.00
	(0.00)	(0.00)
Malaria prevalence	19.51	10.33
	(11.29)	(16.31)
District population	-0.00	-0.00
	(0.00)	(0.00)
Health facility density	0.45	0.38
	(0.30)	(0.37)
TB notification rate	-0.00	0.03
	(0.02)	(0.02)
Nr. of OPD visits	0.00	0.00
	(0.00)	(0.00)
Adjusted R^2	0.32	0.63
AIC	1469.81	1389.47
BIC	1492.37	1412.03
Observations	317	317

Cluster robust standard errors in parentheses; * < 0.05, ** < 0.01, *** < 0.001

Notes: (1) Index variables for maintenance of storage and dispensing area results from the sum of questions listedin appendix A5. (2) The estimation is performed using ordinary least squares.

Figure 2: Trends in the share of children under five sleeping under ITNs between 2007 and 2011 and malaria incidence between 2005 and 2010. Source: DHS (2005-2011).



Note: 95% confidence intervals in grey.

or below, with no statistically significant difference in usage.

Second, it is reasonable to assume that the usage of medicines and thus their availability should be somewhat related to incidence of different diseases. Malaria is among the diseases that got more attention in sub-Saharan Africa and elsewhere. Tanzania experienced large efforts to reduce morbidity and mortality with nation-wide initiatives, such as the Roll Back Malaria Partnership. The right box in figure 2 shows cluster data on incidence of malaria, suggesting little difference in trends between control and treatment groups over the period 2005-2010.

4.2.2. Alternative model specifications

To check the consistency of our results, we ran alternative model specifications with the same data. First, we estimated the regression model using a panel data specification with health facility fixed effects. Second, we ran the same model with fixed effects only on the balanced panel represented by the sub-sample of 91 health facilities observed over the two waves (2011 and 2017). Third, we estimated a regression model with a multilevel approach, setting two clustering levels in the data: district and health facility. All results are reported in Table 8.

The results of our robustness checks largely confirm the results in Table 5. The only exception is the multilevel specification that shows a significant effect on antibiotics, other drugs/vaccines and all drugs pooled. Compared to the two panel specifications above, our preferred model appears more conservative both in term of number and magnitude of the significant effects. The multilevel model approach summarized in Table 8 shows point coefficients marginally smaller in absolute value, although identical in magnitude and direction to our main model.

4.2.3. Alternative treatment and control groups

The empirical setup described above is closely connected to the implementation of the Sikika SAM program in the two districts (Kondoa and Mpwapwa) representing our treatment group. The other districts in the Dodoma region represent the control group (Bahi, Chemba, Chamwino, Dodoma Urban and Kongwa). To rule out effects driven by the specific composition of treatment and control groups, we estimated our main model using alternative treatment and control group specifications. The rationale behind our choices is twofold: one the one hand, the Southern control district of Mpwapwa may differ from the Northern control district of Kondoa, for example with respect to health insurance coverage (see Table 1) and political climate around SAM (see appendix A1). On the other hand, the City of Dodoma (capital city of the United Republic of Tanzania) differs largely from the other districts in the region. Accordingly, in Table 9 we

Table 8: Robustness checks; Diff-in-Diff on alternative model specifications

	Number of stock-out days for				
	(1) Antibiotics	(2) Antimalarials	(3) Other drugs and vaccines	(4) Reproductive health	(5) All drugs
Model	Unbalanced panel with facility FE model regression over 2011-2017				
Treatment	94.35	48.07	207.90	-48.70	157.73
	(61.51)	(63.16)	(109.98)	(55.65)	(172.18)
Post	-69.82	29.07	-64.69	59.26	15.83
	(99.88)	(88.89)	(126.86)	(86.62)	(211.71)
Post × Treatment	-87.38***	5.38	-49.41	-31.65	-107.43*
	(25.29)	(15.15)	(32.45)	(19.00)	(51.69)
Other controls			Yes		
Adjusted R^2	0.29	0.46	0.31	0.22	0.27
AIC	3051.24	2812.47	3229.01	2979.67	3519.08
BIC	3096.35	2857.58	3274.12	3024.78	3564.19
Observations	317	317	317	317	317
Model	Balanced panel with facility FE over period 2011-2017				
Treatment	68.65*	26.92	37.07	-22.97*	61.57
	(24.27)	(17.16)	(41.28)	(6.86)	(51.69)
Post	-46.71*	-15.08	-87.03*	-26.64**	-125.13*
	(15.84)	(11.02)	(31.65)	(5.47)	(35.33)
Post × Treatment	-94.03**	-2.15	-64.60	-37.55	-132.19*
	(17.35)	(8.80)	(29.97)	(17.46)	(40.91)
Other controls			Yes		
Adjusted R^2	0.23	0.21	0.19	0.05	0.18
AIC	1994.12	1861.25	2107.21	1943.08	2245.76
BIC	2013.35	1880.47	2126.44	1962.31	2264.99
Observations	182	182	182	182	182
Model	Mixed model multilevel regression over period 2011-2017				
Treatment	60.02***	15.63	43.79	-21.77**	51.26
	(13.69)	(12.45)	(24.72)	(6.99)	(31.95)
Post	-34.31***	-21.60**	-62.33***	-18.32*	-95.30***
	(10.34)	(8.38)	(7.54)	(8.24)	(14.91)
Post × Treatment	-58.83***	3.58	-39.70***	-23.63*	-75.28***
	(8.85)	(6.84)	(8.76)	(10.88)	(16.78)
Other controls			Yes		
AIC	3514.30	3275.64	3674.15	3446.81	3978.14
BIC	3536.86	3301.95	3696.71	3473.12	4004.45
Observations	317	317	317	317	317
Baseline average	70.50	50.20	126.66	48.25	206.13
Baseline standard deviation	71.96	48.49	92.21	63.77	147.92

Cluster robust standard errors in parentheses; * < 0.05, ** < 0.01, *** < 0.001

Notes: (1) Stock-out days computed over the 3 months prior to the survey dates (Sept 2011 for baseline and May 2017 for endline). (2) For balanced and unbalanced panel models, estimation is performed using ordinary least squares. (3) For the unbalanced pane model, the sample includes only health facilities observed in both baseline (2011) and endline (2017) surveys. (4) Multilevel model estimated using Maximum Likelihood (ML) with Expectation-Maximization (EM) algorithm.

report the coefficients of interest for analyses with (1) treatment group including only Mpwapwa district, (2) treatment group including only Kondoa district and (3) a control group that excludes the Dodoma Urban district.

Compared to our main specification in Table 5, the results show point estimates with identical signs. The magnitude of the effects is also similar, with overlapping in confidence intervals that appear wider as a result of the reduction in number of clusters and observations. Under CRSEs, all three alternative models confirm a significant effect on antibiotics. The effect remains significant also for the variable pooling all drugs for the models with alternative treatment group specification, but not for the model excluding the Dodoma Urban district.

4.2.4. Placebo tests

To further support our empirical approach we conducted placebo tests, estimating our DID model on two placebo variables unaffected by the Sikika SAM program. Given the health system focus of our analysis, we focus on three health system variables that had nothing to do with the SAM program, but could have affected our study outcome. First, we look at the share of people exempted from payment at the surveyed facilities. This variable captures both local poverty levels and general health system financing efforts by the government, and should allow us to detect potential differential trends that could bias the results. Second, we look at average yearly rainfall around the treated facilities, which should allow us to identify differences in agricultural productions and incomes across districts. Last, we look at the share of households reporting children under five sleeping under ITNs, which we consider both a proxy for (vertical) program-specific efforts by the government and a potential proxy for the under-5 patient burden at facilities.

We find no statistically significant associations between the SAM program and any of the three placebo outcomes (Table 10).

5. Conclusion

Our study analysed data from a survey on government-managed health facilities in the Tanzanian region of Dodoma. We evaluated the effect of a specific social accountability monitoring program implemented by the local NGO Sikika in two treatment districts out of the seven composing the region. The outcomes evaluated - related to the SAM program objectives – are duration of stock-outs in essential medicines and infrastructure maintenance. We employed a difference-in-difference approach and included a series of control variables to account for many factors that potentially affect the outcomes of interest. Additionally, the analytical strategy accounted for the small number of clusters that make the standard sandwich estimator for standard errors inefficient.

Concerning the availability of essential medicines, we found the SAM impact to be significant only for antibiotics and the pool of all drugs considered. The lack of effect for antimalarials and medicines and commodities prescribed in reproductive health can be interpreted in light of the distribution policy in place in Tanzania. In fact, the lack of significant association between SAM and stock-out days of drugs against malaria in our models may be related to the coverage of the disease by large multilateral initiatives and vertical programs, whose supply-chains operate outside of the government channels. Likewise, in the context of a wider RCHS program, central government authorities are committed to directly "plan, procure and distribute reproductive health commodities including those for family planning at the council level countrywide". All other categories of drugs - including antibiotics and other generic medicines/vaccines – are instead procured and distributed to government-managed health facilities through the regular supply-chain, that is the mechanisms that expect to be influenced by the SAM program.

Unfortunately, our analysis could not detect any effect on the other dimensions of health facility performance addressed by the SAM program. The analysis ran on the indexes of infrastructure investments and

Table 9: Robustness checks; Diff-in-Diff specification with alternative control and treatment groups

	Number of stock-out days for				
	(1) Antibiotics	(2) Antimalarials	(3) Other drugs and vaccines	(4) Reproductive health	(5) All drugs
Treatment group	Only Mpwapwa				
Control group	Bahi, Chemba, Chamwino, Dodoma, Kondoa, Kongwa				
Treatment	61.36*	23.95	79.16*	3.63	106.35
	(22.61)	(19.04)	(29.87)	(16.60)	(47.20)
Post	-17.40	-11.02	-40.24	-4.07	-54.59
	(17.11)	(18.38)	(21.86)	(18.26)	(41.33)
Post × Treatment	-46.15*	5.51	-38.71*	-31.21	-75.17**
	(14.53)	(7.89)	(14.23)	(17.25)	(18.74)
Other controls	Yes				
Adjusted R^2	0.13	0.15	0.16	0.02	0.09
AIC	3526.05	3275.39	3685.82	3446.66	3979.62
BIC	3548.60	3297.94	3708.37	3469.21	4002.18
Observations	317	317	317	317	317
Treatment group	Only Kondoa				
Control group	Bahi, Chemba, Chamwino, Dodoma, Kongwa, Mpwapwa				
Treatment	50.61	4.72	5.80	-44.86*	-14.61
	(22.22)	(20.99)	(31.66)	(17.21)	(49.98)
Post	-79.82**	-31.30**	-104.64**	5.22	-135.04***
	(17.23)	(8.14)	(20.37)	(8.79)	(22.65)
Post × Treatment	-50.74**	7.06	-31.43	-20.86**	-60.42*
	(9.54)	(8.21)	(13.61)	(5.18)	(16.33)
Other controls	Yes				
Adjusted R^2	0.14	0.16	0.17	0.02	0.10
AIC	3519.86	3273.32	3679.59	3446.63	3975.19
BIC	3542.41	3295.88	3702.14	3469.19	3997.74
Observations	317	317	317	317	317
Treatment group	Both intervention districts (Kondoa and Mpwapwa)				
Control group	Bahi, Chemba, Chamwino, Kongwa				
Treatment	81.22**	28.87	75.12*	-6.42	104.66*
	(17.29)	(16.70)	(22.13)	(12.36)	(36.31)
Post	-41.79**	-28.76*	-93.53**	-14.64*	-126.33***
	(9.49)	(9.08)	(17.36)	(5.49)	(17.14)
Post × Treatment	-55.58**	6.89	-37.52	-25.90	-73.56
	(11.98)	(7.03)	(26.23)	(13.14)	(36.51)
Other controls	Yes				
Adjusted R^2	0.17	0.20	0.19	0.01	0.12
AIC	3163.91	2930.43	3307.42	3101.40	3565.36
BIC	3182.17	2948.69	3325.68	3119.66	3583.62
Observations	285	285	285	285	285
Baseline average	70.50	50.20	126.66	48.25	206.13
Baseline standard deviation	71.96	48.49	92.21	63.77	147.92

Cluster robust standard errors in parentheses; * < 0.05, ** < 0.01, *** < 0.001

Notes: (1) The estimation is performed using ordinary least squares. (2) Stock-out days are computed over the 3 months prior to the survey dates (September 2011 for baseline and May 2017 for endline).

Table 10: Placebo tests; Diff-in-Diff specification for three Placebo variables over period 2011-2017

	(1)	(2)	(3)
	Average yearly rainfall	ITN usage	Exempted patients
Treatment	30.45	0.17	-1.24
	(62.12)	(0.13)	(0.86)
Post	61.67	-0.11	-2.83***
	(71.77)	(0.11)	(0.38)
Post × Treatment	-12.78	-0.10	0.55
	(42.54)	(0.10)	(0.97)
Health centre	-6.91	-0.00	0.20
	(12.36)	(0.01)	(0.30)
District hospital	11.20	0.05	-0.01
	(11.13)	(0.07)	(0.33)
Active HF committee	9.76	-0.02	0.75
	(11.37)	(0.02)	(0.85)
Distance from MSD store	0.00	-0.00	0.00
	(0.00)	(0.00)	(0.00)
Urban area	1.01	-0.34**	-1.06
	(30.08)	(0.09)	(0.87)
Share of pop. in 1st wealth quintile	-135.87**	-0.08	-1.89
	(26.79)	(0.05)	(1.26)
Share of pop. in 5st wealth quintile	38.10	0.49*	0.44
	(50.09)	(0.17)	(1.50)
Malaria prevalence	-815.32	2.27	-24.46
	(637.00)	(1.77)	(13.39)
District population	-0.00	-0.00	0.00***
	(0.00)	(0.00)	(0.00)
Health facility density	-5.87	-0.02	-0.07
	(11.22)	(0.04)	(0.54)
TB notification rate	1.07	0.00	-0.00
	(0.78)	(0.00)	(0.00)
Nr. of OPD visits	0.01	0.00	-0.00
	(0.01)	(0.00)	(0.00)
Average yearly rainfall		0.00	0.00
		(0.00)	(0.00)
Adjusted R^2	0.44	0.56	
Pseudo R^2			0.08
AIC	3618.94	-337.51	887.87
BIC	3641.49	-314.96	909.87
Observations	317	317	289

Cluster robust standard errors in parentheses; * < 0.05, ** < 0.01, *** < 0.001

Notes: (1) The full Diff-in-Diff specification includes the Average yearly rainfall. The variable was excluded in the analysis in column (1). (2) The estimation is performed using ordinary least squares.

maintenance showed no significant effect, whilst we did not have reliable data for financial management, utilization of funds at local level and functioning of health facility committees. This clearly limits the scope of the measured impact of Sikika's program and the general conclusions to be drawn. The SAM program considered here is likely to exert an effect on performance domains related to providers' effort, directly affected by the increased accountability generated at service delivery level. Other performance domains involving accountability at higher level (eg. district budgeting processes) or changes in management systems seemed to be unaffected and probably require a different approach to accountability. To this extent, future research should explore the impact of social accountability on other structural and financial outcomes in the domain of health and health care, as well as the long run and spill over effects on the culture of accountability within district authorities and health staff.

Our study has some relevant limitations. First, our analysis is based only on two time points, i.e. baseline and endline. The analysis would have benefited from the availability of pre-intervention time points, and would also have benefited from more intermediate outcomes to assess the quality of implementation of both the HPSS and the Sikika SAM programs. Second, despite the favorable setup provided by the Dodoma region, we exploited a natural experiment that did not allow us to control for all possible confounders. We are aware that some factors excluded from our analysis may exert an influence either on our outcomes of interest or on the social accountability mechanism that we are evaluating. Our main assumption is that these factors did not change in way systematically correlated with the program assignments. In the absence of cluster-randomized controlled trials (RCTs) in this area future research would benefit from an earlier involvement between researcher and implementation projects such as HPSS, in order to plan a solid data collection strategy with relevant control variables and adequate pre-treatment and intermediate outcome (Leatherdale, 2019). Last, the very nature of our study limits the external validity of the results, a feature shared with many other existing studies assessing impact of social accountability (McCoy et al., 2012; Molina et al., 2017; O'Meally, 2013).

The results presented in this paper support the idea that well designed and properly embedded SAM approaches can be effective in improving selected aspects of health systems performance, and may also generate wider and more long term benefits by creating a culture of accountability in the communities involved. A key element that we must acknowledge in reading the results is the positive and receptive environment among district authorities favoured by the region-wide implementation of the HPSS project. Our analysis of the mechanism that generated the measured impact on availability of essential medicines builds on the context. First, Sikika is a well-known "watchdog" organization that partnered with local grassroots level CSOs in the districts where the SAM was implemented, providing accurate knowledge of the social setting and generating a sense of recognisability within the community. Second, the SAM process seemed to be designed according to guidelines and best practices found in the literature, notably involving equally the supply-side (district authorities, village/ward leaders) and the demand-side (community representatives, CSOs, media and other stakeholders), putting the two together and providing feedback to both. In this context, our hypothesis is that Sikika reduced the distance between the existing institutionalized social accountability bodies (HFGCs) and the district authorities, thus strengthening the accountability relationship.

The features above are consistent with the previous literature on social accountability initiatives. For example, several authors recognized the importance of identifying and evaluating the context within which the social accountability initiatives are implemented (Joshi, 2013; Westhorp et al., 2014). To this extent, a useful framework is provided by Lodenstein et al. (2017b), that explain how the social accountability process itself is directly influenced by external and contextual factors such as wider social/cultural/political/economic context or community and provider characteristics. These external factors can either underpin or diminish the effect of the social accountability program and should thus be addressed appropriately. Other authors isolated a set of common contextual characteristics for successful social accountability initiatives, such as: community participation and feedback mechanisms in the context of health service decentralization, provision of adequate performance information to the community, involvement of existing NGOs and CSOs,

working outside of bureaucratic mechanisms, build on existing productive relations and leverage on willing political leaders (Berlan and Shiffman, 2012; Björkman Nyqvist et al., 2017; Cleary et al., 2013; Danhoundo et al., 2018; Dewachter et al., 2018; Fox, 2015; O'Meally, 2013; Ringold et al., 2012; Schaaf et al., 2017).

Our study has important implications for LMICs planning to introduce social accountability mechanisms as tools to improve bottom-up governance in the health system. First, the body of evidence available, above and beyond the results presented here, should be used more consistently to inform decisions and shape policies. In our specific case, the lack of effect on outcomes related to infrastructure maintenance should suggest changes in the structure of the SAM program or the scope of the program itself. Second, our study proposes a structured attempt of evaluation for all policies or pilot projects across LMICs that do not build on a robust data structure for assessment. Despite the great push for RCT-based evidence in development policies across LMICs in the last decades and the related debate (Barrett and Carter, 2010), the number of unexplored opportunities for valuable non-randomised evaluations (for example in the context of operational research) remains very large, and so is the unexpressed potential for improvements in decisions and policies. On a final note, social accountability initiatives intersect with another salient topic in the debate around universal health coverage in Tanzania: the community health fund (CHF). As one can infer from its name, the CHF is a community-based health insurance scheme employed to improve financial health protections among rural and poor urban communities. Social accountability initiatives play an important role in determining the success of the CHF schemes, favoring acceptance and ownership within the local community (Kalolo et al., 2015; Kigume and Maluka, 2018; Pedersen and Jacob, 2018; Renggli et al., 2019).

Conflicts of interest

The authors report no conflicts of interest.

References

Adhvaryu, A., Nyshadham, A., 2015. Return to Treatment in the Formal Health Care Sector: Evidence from Tanzania. American Economic Journal: Economic Policy 7, 29–57. doi:10.1257/pol.20120262.

Angrist, J.D., Pischke, J.S., 2009. Mostly harmless econometrics: an empiricist's companion. Princeton University Press, Princeton, USA.

Attaei, M.W., Khatib, R., McKee, M., Lear, S., Dagenais, G., Igumbor, E.U., AlHabib, K.F., Kaur, M., Kruger, L., Teo, K., Lanas, F., Yusoff, K., Oguz, A., Gupta, R., Yusufali, A.M., Bahonar, A., Kutty, R., Rosengren, A., Mohan, V., Avezum, A., Yusuf, R., Szuba, A., Rangarajan, S., Chow, C., Yusuf, S., 2017. Availability and affordability of blood pressure-lowering medicines and the effect on blood pressure control in high-income, middle-income, and low-income countries: an analysis of the pure study data. The Lancet Public Health 2, e411–e419. doi:10.1016/S2468-2667(17)30141-X.

Barrett, C.B., Carter, M.R., 2010. The Power and Pitfalls of Experiments in Development Economics: Some Non-random Reflections. Applied Economic Perspectives and Policy 32, 515–548. doi:10.1093/aepp/ppq023.

Barrington, J., Wereko-Brobby, O., Ward, P., Mwafongo, W., Kungulwe, S., 2010. SMS for Life: a pilot project to improve anti-malarial drug supply management in rural Tanzania using standard technology. Malaria Journal 9, 298. doi:10.1186/1475-2875-9-298.

Bate, R., Hess, K., Mooney, L., 2010. Antimalarial medicine diversion: stock-outs and other public health problems. Research and Reports in Tropical Medicine , 19doi:10.2147/RRTM.S13242.

Berheto, T.M., Haile, D.B., Mohammed, S., 2014. Predictors of Loss to follow-up in Patients Living with HIV/AIDS after Initiation of Antiretroviral Therapy. North American Journal of Medical Sciences 6, 453–459. doi:10.4103/1947-2714.141636.

Berlan, D., Shiffman, J., 2012. Holding health providers in developing countries accountable to consumers: a synthesis of relevant scholarship. Health Policy and Planning 27, 271–280. doi:10.1093/heapol/czr036.

Bigdeli, M., Jacobs, B., Tomson, G., Laing, R., Ghaffar, A., Dujardin, B., Van Damme, W., 2013. Access to medicines from a health system perspective. Health Policy and Planning 28, 692–704. doi:10.1093/heapol/czs108.

Björkman, M., Svensson, J., 2009. Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda. The Quarterly Journal of Economics 124, 735–769. doi:10.1162/qjec.2009.124.2.735.

Björkman Nyqvist, M., de Walque, D., Svensson, J., 2017. Experimental Evidence on the Long-Run Impact of Community-Based Monitoring. American Economic Journal: Applied Economics 9, 33–69. doi:10.1257/app.20150027.

Cameron, A., Ewen, M., Ross-Degnan, D., Ball, D., Laing, R., 2009. Medicine prices, availability, and affordability in 36 developing and middle-income countries: a secondary analysis. The Lancet 373, 240–249. doi:10.1016/S0140-6736(08)61762-6.

Cameron, A.C., Gelbach, J., Miller, D., 2008. Bootstrap-Based Improvements for Inference with Clustered Errors. The Review of Economics and Statistics 90, 414–427. doi:10.1162/rest.90.3.414.

Cleary, S.M., Molyneux, S., Gilson, L., 2013. Resources, attitudes and culture: an understanding of the factors that influence the functioning of accountability mechanisms in primary health care settings. BMC health services research 13, 320. doi:10.1186/1472-6963-13-320.

Danhoundo, G., Nasiri, K., Wiktorowicz, M.E., 2018. Improving social accountability processes in the health sector in sub-Saharan Africa: a systematic review. BMC Public Health 18, 497. doi:10.1186/s12889-018-5407-8.

De Savigny, D., Adam, T. (Eds.), 2009. Systems thinking for health systems strengthening. Alliance for Health Policy and Systems Research ; World Health Organization, Geneva.

Dewachter, S., Holvoet, N., Kuppens, M., Molenaers, N., 2018. Beyond the Short versus Long Accountability Route Dichotomy: Using Multi-track Accountability Pathways to Study Performance of Rural Water Services in Uganda. World Development 102, 158–169. doi:10.1016/j.worlddev.2017.09.018.

Dimick, J.B., Ryan, A.M., 2014. Methods for Evaluating Changes in Health Care Policy: The Difference-in-Differences Approach. JAMA 312, 2401–2402. doi:10.1001/jama.2014.16153.

Donald, S.G., Lang, K., 2007. Inference with Difference-in-Differences and Other Panel Data. The Review of Economics and Statistics 89, 221–233. doi:10.1162/rest.89.2.221.

Eze, I.C., Kramer, K., Msengwa, A., Mandike, R., Lengeler, C., 2014. Mass distribution of free insecticide-treated nets do not interfere with continuous net distribution in Tanzania. Malaria Journal 13, 196. doi:10.1186/1475-2875-13-196.

Fadlallah, R., El-Jardali, F., Hemadi, N., Morsi, R.Z., Abou Samra, C.A., Ahmad, A., Arif, K., Hishi, L., Honein-AbouHaidar, G., Akl, E.A., 2018. Barriers and facilitators to implementation, uptake and sustainability of community-based health insurance schemes in low- and middle-income countries: a systematic review. International Journal for Equity in Health 17, 13. doi:10.1186/s12939-018-0721-4.

Favin, M., Steinglass, R., Fields, R., Banerjee, K., Sawhney, M., 2012. Why children are not vaccinated: a review of the grey literature. International Health 4, 229–238. doi:10.1016/j.inhe.2012.07.004.

Feruglio, F., Nisbett, N., 2018. The challenges of institutionalizing community-level social accountability mechanisms for health and nutrition: a qualitative study in Odisha, India. BMC Health Services Research 18, 788. doi:10.1186/s12913-018-3600-1.

Fiala, N., Premand, P., 2018. Social Accountability and Service Delivery: Experimental Evidence from Uganda doi:10.4419/86788874.

Fox, J., 2007. The uncertain relationship between transparency and accountability. Development in Practice 17, 663–671. doi:10.1080/09614520701469955.

Fox, J., 2015. Social Accountability: What Does the Evidence Really Say? World Development 72, 346–361. doi:10.1016/j.worlddev.2015.03.011.

Frumence, G., Nyamhanga, T., Mwangu, M., Hurtig, A.K., 2014. Participation in health planning in a decentralised health system: Experiences from facility governing committees in the Kongwa district of Tanzania. Global Public Health 9, 1125–1138. doi:10.1080/17441692.2014.953563.

Gaventa, J., McGee, R., 2013. The Impact of Transparency and Accountability Initiatives. Development Policy Review 31, s3–s28. doi:10.1111/dpr.12017.

Githinji, S., Kigen, S., Memusi, D., Nyandigisi, A., Mbithi, A.M., Wamari, A., Muturi, A.N., Jagoe, G., Barrington, J., Snow, R.W., Zurovac, D., 2013. Reducing Stock-Outs of Life Saving Malaria Commodities Using Mobile Phone Text-Messaging: SMS for Life Study in Kenya. PLOS ONE 8, e54066. doi:10.1371/journal.pone.0054066.

Greer, S.L., Wismar, M., Figueras, J. (Eds.), 2016. Strengthening health system governance: better policies, stronger performance. Open University Press, Maidenhead, UK.

Hamal, M., de Cock Buning, T., De Brouwere, V., Bardají, A., Dieleman, M., 2018. How does social accountability contribute to better maternal health outcomes? A qualitative study on perceived changes with government and civil society actors in Gujarat, India. BMC Health Services Research 18, 653. doi:10.1186/s12913-018-3453-7.

Ikoh, M.U., Udo, A.U., Charles, A.O., Charles, J.O., 2009. The Influence of "Stock out" on Health-Seeking Behavior of Low Income Women in Uyo Urban, Akwa Ibom State, Nigeria. International Quarterly of Community Health Education 29, 257–273. doi:10.2190/IQ.29.3.e.

Joshi, A., 2013. Do They Work? Assessing the Impact of Transparency and Accountability Initiatives in Service Delivery. Development Policy Review 31, s29–s48. doi:10.1111/dpr.12018.

Joshi, A., 2017. Legal Empowerment and Social Accountability: Complementary Strategies Toward Rights-based Development in Health? World Development 99, 160–172. doi:10.1016/j.worlddev.2017.07.008.

Joshi, A., Houtzager, P.P., 2012. Widgets or Watchdogs? Public Management Review 14, 145–162. doi:10.1080/14719037.2012.657837.

Kalolo, A., Radermacher, R., Stoermer, M., Meshack, M., Allegri, M.D., 2015. Factors affecting adoption, implementation fidelity, and sustainability of the Redesigned Community Health Fund in Tanzania: a mixed methods protocol for process evaluation in the Dodoma region. Global Health Action 8, 29648. doi:10.3402/gha.v8.29648.

Kamuzora, P., Gilson, L., 2007. Factors influencing implementation of the Community Health Fund in Tanzania. Health Policy and Planning 22, 95–102. doi:10.1093/heapol/czm001.

Kessy, F.L., 2014. Improving Health Services through Community Participation in Health Governance Structures in Tanzania. Journal of Rural & Community Development 9.

Kigume, R., Maluka, S., 2018. Health sector decentralisation in Tanzania: How do institutional capacities influence use of decision space? The International Journal of Health Planning and Management 33, e1050–e1066. doi:10.1002/hpm.2587.

Kruk, M.E., Rockers, P.C., Mbaruku, G., Paczkowski, M.M., Galea, S., 2010. Community and health system factors associated with facility delivery in rural Tanzania: A multilevel analysis. Health Policy 97, 209–216. doi:10.1016/j.healthpol.2010.05.002.

Layer, E.H., Kennedy, C.E., Beckham, S.W., Mbwambo, J.K., Likindikoki, S., Davis, W.W., Kerrigan, D.L., Brahmbhatt, H., Team, t.L.T.C.S., 2014. Multi-Level Factors Affecting Entry into and Engagement in the HIV Continuum of Care in Iringa, Tanzania. PLOS ONE 9, e104961. doi:10.1371/journal.pone.0104961.

Leatherdale, S.T., 2019. Natural experiment methodology for research: a review of how different methods can support real-world research. International Journal of Social Research Methodology 22, 19–35. doi:10.1080/13645579.2018.1488449.

Leung, N.H.Z., Chen, A., Yadav, P., Gallien, J., 2016. The Impact of Inventory Management on Stock-Outs of Essential Drugs in Sub-Saharan Africa: Secondary Analysis of a Field Experiment in Zambia. PLOS ONE 11, e0156026. doi:10.1371/journal.pone.0156026.

Lewis, M., 2006. Governance and Corruption in Public Health Care Systems. SSRN Scholarly Paper ID 984046. Social Science Research Network. Rochester, NY.

Lewis, M., Pettersson, G., 2009. Governance In Health Care Delivery : Raising Performance. Policy Research Working Papers, The World Bank. doi:10.1596/1813-9450-5074.

Lodenstein, E., Dieleman, M., Gerretsen, B., Broerse, J.E.W., 2017a. Health provider responsiveness to social accountability initiatives in low- and middle-income countries: a realist review. Health Policy and Planning 32, 125–140. doi:10.1093/heapol/czw089.

Lodenstein, E., Ingemann, C., Molenaar, J.M., Dieleman, M., Broerse, J.E.W., 2018. Informal social accountability in maternal health service delivery: A study in Northern Malawi. PLOS ONE 13, e0195671. doi:10.1371/journal.pone.0195671.

Lodenstein, E., Mafuta, E., Kpatchavi, A.C., Servais, J., Dieleman, M., Broerse, J.E.W., Barry, A.A.B., Mambu, T.M.N., Toonen, J., 2017b. Social accountability in primary health care in West and Central Africa: exploring the role of health facility committees. BMC Health Services Research 17, 403. doi:10.1186/s12913-017-2344-7.

Macha, J., Mushi, H., Borghi, J., 2011. Examining the links between accountability, trust and performance in health service delivery in Tanzania. Ifakara Health Institute: Dar es Salaam .

MacKinnon, J.G., Webb, M.D., 2018. The wild bootstrap for few (treated) clusters. The Econometrics Journal 21, 114–135. doi:10.1111/ectj.12107.

Martin Hilber, A., Blake, C., Bohle, L.F., Bandali, S., Agbon, E., Hulton, L., 2016. Strengthening accountability for improved maternal and newborn health: A mapping of studies in Sub-Saharan Africa. International Journal of Gynecology & Obstetrics 135, 345–357. doi:10.1016/j.ijgo.2016.09.008.

McCoy, D.C., Hall, J.A., Ridge, M., 2012. A systematic review of the literature for evidence on health facility committees in low- and middle-income countries. Health Policy and Planning 27, 449–466. doi:10.1093/heapol/czr077.

McNeil, M., Male, C. (Eds.), 2010. Demanding Good Governance: Lessons from Social Accountability Initiatives in Africa. The World Bank. doi:10.1596/978-0-8213-8380-3.

Mikkelsen-Lopez, I., Shango, W., Barrington, J., Ziegler, R., Smith, T., deSavigny, D., 2014. The challenge to avoid anti-malarial medicine stock-outs in an era of funding partners: the case of Tanzania. Malaria Journal 13, 181. doi:10.1186/1475-2875-13-181.

Ministry of Health, 2002. Baseline Survey of the Pharmaceutical Sector in Tanzania 2002. Technical Report. United Republic of Tanzania. Dodoma.

Ministry of Health and Social Welfare [MoHSW], 2011. Annual Health Statistical Tables and Figures Tanzania Mainland. HMIS 2011. Technical Report. United Republic of Tanzania. Dar es Salaam, Tanzania.

Ministry of Health, Community Development, Gender, Elderly and Children [MoHCDGEC], 2019. Health Facility Registry (HFR).

MoHSW, Ifakara Health Institute, 2013. Tanzania - Service Availability and Readiness Assessment (SARA) 2012. Technical Report. United Republic of Tanzania. Dar es Salaam, Tanzania.

MoHSW, Ministry of Health, National Bureau of Statistics, Office of Chief Government Statistician, ICF International, 2016. Tanzania Service Provision Assessment Survey 2014-2015. Technical Report. United Republic of Tanzania and ICF International. Dar es Salaam, Tanzania.

Molina, E., Carella, L., Pacheco, A., Cruces, G., Gasparini, L., 2017. Community monitoring interventions to curb corruption and increase access and quality in service delivery: a systematic review. Journal of Development Effectiveness 9, 462–499. doi:10.1080/19439342.2017.1378243.

Moïsi, J.C., Gatakaa, H., Noor, A.M., Williams, T.N., Bauni, E., Tsofa, B., Levine, O.S., Scott, J.A.G., 2010. Geographic access to care is not a determinant of child mortality in a rural Kenyan setting with high health facility density. BMC Public Health 10, 142. doi:10.1186/1471-2458-10-142.

Muhamadi, L., Nsabagasani, X., Tumwesigye, M.N., Wabwire-Mangen, F., Ekström, A.M., Peterson, S., Pariyo, G., 2010. Inadequate pre-antiretroviral care, stock-out of antiretroviral drugs and stigma: Policy challenges/bottlenecks to the new WHO recommendations for earlier initiation of antiretroviral therapy (cd¡350 cells/microl) in eastern Uganda. Health Policy 97, 187–194. doi:10.1016/j.healthpol.2010.06.003.

Mujinja, P.G.M., Kida, T.M., 2014. Implications of health sector reforms in Tanzania: policies, indicators and accessibility to health services. Economic and Social Research Foundation, Dar es Salaam, Tanzania.

Musau, S., Chee, G., Patsika, R., Malangalila, E., Chitama, D., Van Praag, E., Schettler, G., 2011. Tanzania Health System Assessment 2010 Report. Technical Report. Health Systems 20/20 project, Abt Associates Inc.. Bethesda, MD.

National Bureau of Statistics, Macro International Inc., 2007. Tanzania Service Provision Assessment Survey 2006. Technical Report. National Bureau of Statistics and Macro International Inc.. Dar es Salaam, Tanzania.

National Bureau of Statistics, Office of Chief Government Statistician, 2013. 2012 Popolation and Housing Census. Population Distribution by Administrative Areas. Technical Report. United Republic of Tanzania. Dar es Salaam, Tanzania.

Ogbuabor, D.C., Onwujekwe, O.E., 2018. The community is just a small circle: citizen participation in the free maternal and child healthcare programme of Enugu State, Nigeria. Global Health Action 11, 1421002. doi:10.1080/16549716.2017.1421002.

O'Meally, S.C., 2013. Mapping Context for Social Accountability: A Resource Paper. Technical Report 79351. World Bank. Washington DC, USA.

Pedersen, R.H., Jacob, T., 2018. Social protection in an electorally competitive environment (2): The politics of health insurance in Tanzania. Technical Report 110. ESID, The University of Manchester. Manchester, UK.

Renggli, S., Mayumana, I., Mshana, C., Mboya, D., Kessy, F., Tediosi, F., Pfeiffer, C., Aerts, A., Lengeler, C., 2019. Looking at the bigger picture: how the wider health financing context affects the implementation of the Tanzanian Community Health Funds. Health Policy and Planning doi:10.1093/heapol/czy091.

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

Rokicki, S., Cohen, J., Fink, G., Salomon, J.A., Landrum, M.B., 2018. Inference With Difference-in-Differences With a Small Number of Groups: A Review, Simulation Study, and Empirical Application Using SHARE Data. Medical Care 56, 97–105. doi:10.1097/MLR.0000000000000830.

Roodman, D.M., MacKinnon, J.G., Nielsen, M.r., Webb, M., 2018. Fast and Wild: Bootstrap Inference in Stata Using boottest. Working Paper 1406. Queen's Economics Department Working Paper.

Schaaf, M., Topp, S.M., Ngulube, M., 2017. From favours to entitlements: community voice and action and health service quality in Zambia. Health Policy and Planning 32, 847–859. doi:10.1093/heapol/czx024.

Sikika, 2013. Annual Report 2012. Technical Report. Sikika. Dodoma, Tanzania.

Sikika, 2016. Ripoti ya Ufuatiliaji na Uwajibikaji kwa Jamii (SAM). Halmashauri ya Wilaya ya Kondoa Vijijini. Imeandaliwa na Timu ya SAM. Septemba 2016. Technical Report. Sikika. Dodoma, Tanzania.

Stoermer, M., 2017. JAZIA Prime Vendor System Dodoma Region, Tanzania. Technical Report. Swiss Tropical and Public Health Institute (Swiss TPH). Basel, Switzerland.

Tefera, W., Tesfaye, H., Bekele, A., Kayessa, E., Waltensperger, K.Z., Marsh, D.R., 2014. Factors influencing the low utilization of curative child health services in Shebedino District, Sidama Zone, Ethiopia. Ethiopian Medical Journal 52 Suppl 3, 109–117.

United Nations (Ed.), 2012. The global partnership for development: making rhetoric a reality. 2012, United Nations, New York, USA.

University Consultancy Bureau, 2018. Health Facility Survey Report for A Follow Up Survey for the Health Promotion and System Strengthening (HPSS). Technical Report. University of Dar es Salaam. Dar es Salaam, Tanzania.

Wagenaar, B.H., Gimbel, S., Hoek, R., Pfeiffer, J., Michel, C., Manuel, J.L., Cuembelo, F., Quembo, T., Afonso, P., Gloyd, S., Sherr, K., 2014. Stock-outs of essential health products in Mozambique – longitudinal analyses from 2011 to 2013. Tropical Medicine & International Health 19. doi:10.1111/tmi.12314.

Wales, J., Tobias, J., Malangalila, E., Swai, G., Wild, L., 2014. Stock-outs of essential medicines in Tanzania. A political economy approach to analysing problems and identifying solutions. Technical Report. Overseas Development Institute. London, UK.

Webb, M.D., 2013. Reworking Wild Bootstrap Based Inference for Clustered Errors. Working Paper 1315.

Westhorp, G., Walker, B., Rogers, P., Overbeeke, N., Ball, D., Brice, G., 2014. Enhancing community accountability, empowerment and education outcomes in low and middle-income countries: a realist review. EPPI-Centre, Social Science Research Unit, Institute of Education, University of London, London, UK.

Wiedenmayer, K., Mbwasi, R., Mfuko, W., Mpuya, E., Charles, J., Chilunda, F., Mbepera, D., Kapologwe, N., 2019. Jazia prime vendor system- a public-private partnership to improve medicine availability in Tanzania: from pilot to scale. Journal of Pharmaceutical Policy and Practice 12. doi:10.1186/s40545-019-0163-4.

Wooldridge, J.M., 2010. Econometric analysis of cross section and panel data. 2nd ed ed., MIT Press, Cambridge, USA.

Appendix A.

802 *Appendix A.1. Implementation timeline for HPSS project and SAM program in Dodoma*

Year	Month	Actor	Event
2011	August-September	HPSS	Baseline survey in Dodoma region
2012	October	Sikika	First SAM round starts in Kondoa and Mpwapwa districts
2014	February	Sikika	Second SAM round starts in Kondoa and Mpwapwa districts
2014	July	Sikika	District council decides to stop Sikika activities. Sikika reacts in court
2014	August-December	Sikika	The media brings the events in Kondoa to the attention
2014	September	HPSS	Jazia Prime vendor (PV) system starts operations to address stock outs
2014	November	HPSS	Kondoa district receives first orders through Jazia PV
2015	January-June	HPSS	Mpwapwa district receives first orders through Jazia PV
2015	January-December	Sikika	Follow up and monitoring activities in Mpwapwa
2015	May-November	Sikika	Hearings before the Tanzania High Court (Kondoa case)
2015	June	HPSS	Health facilities in all districts have placed orders with Jazia PV
2015	December	Sikika	District council removes ban on Sikika activities
2016	January-December	Sikika	Follow up and monitoring activities in Mpwapwa
2016	August	Sikika	Third SAM round starts in Kondoa, after the ban
2017	January-June	Sikika	Follow up and monitoring activities in Kondoa and Mpwapwa
2017	May-June	HPSS	Endline survey in Dodoma region

Sources: Sikika (2013, 2014, 2015, 2016); Stoermer (2017); The Citizen (2014).

Appendix A.2. Composition of control and treatment groups across districts over time

CONTROL group

	Kongwa (*count and %*)	Dodoma Urban (*count and %*)	Bahi (*count and %*)	Chamwino (*count and %*)	Chemba (*count and %*)	Total (*count and %*)
Pre	32	19	24	49	22	146
	21.92	13.01	16.44	33.56	15.07	100.00
Post	18	13	15	21	16	83
	21.69	15.66	18.07	25.30	19.28	100.00
Total	50	32	39	70	38	229
	21.83	13.97	17.03	30.57	16.59	100.00

TREATMENT group

	Kondoa (*count and %*)	Mpwapwa (*count and %*)	Total (*count and %*)
Pre	22	37	59
	37.29	62.71	100.00
Post	9	20	29
	31.03	68.97	100.00
Total	31	57	88
	35.23	64.77	100.00

Appendix A.3. Data sources for Control Variables

Variable	Level	Source of data	Years
Active HF committee (Yes/No)	Health facility	HPSS base- and endline surveys	2011,2017
Urban area (Yes/No)	Cluster	AIS 2011 and MIS 2017	2011, 2017
Share of pop. wealth index Q1/Q5	Cluster	DHS 2010, AIS 2011 and MIS 2017	2011, 2017
District population (Number)	District	Annual Health Statistics, MoHSW	2011, 2017
Health facility density (per 100'000)	District	Annual Health Statistics, MoHSW	2011, 2017
TB notification rate (per 100'000)	District	Annual Health Statistics, MoHSW	2011, 2017
Malaria prevalence (Parasite rate)	Cluster	DHS Covariates Extract Data	2010, 2015
Nr. of OPD visits (Number)	District	Annual Health Statistics, MoHSW	2011, 2017
Average yearly rainfall (mm per year)	Cluster	DHS Covariates Extract Data	2010, 2015
Distance from MSD store (km)	Health facility	GIS Open data Esri and NBS	2011, 2017
Health insurance coverage	Household	DHS 2010 and DHS 2011	2010, 2011
Children sleeping under bednet	Household	DHS 2010 and DHS 2011	2010, 2011
Health facility delivery	Individual	DHS 2010	2010
Children birthweight	Individual	DHS 2010	2010
Anemic children	Individual	DHS 2010	2010
Children vaccination	Individual	DHS 2010	2010

Notes: (1) Cluster characteristics are matched with health facility based on shortest distance, using QGIS software for spatial analysis. (2) Annual Health Statistics produced by National Bureau of Statistics (NBS) and Ministry of Health and Social Welfare (MoHSW) of Tanzania. As of 2019, the name changed to Ministry of Health, Community Development, Gender, Elderly and Children (MoHCDGEC). (3) Distance from MSD store computed using GIS Open Data produced by Esri Eastern Africa using QGIS software for spatial analysis.

Appendix A.4. Questions composing the index of infrastructure maintenance of medicines storage and dispensing areas

The following questions were asked as part of all facility assessments for infrastructure maintenance (storage room and the dispensing area) conducted at baseline and endline. The range of allowed answers was limited ot "Yes" or "No". Source: University Consultancy Bureau (2018).

1. There is a method in place to control temperature (e.g. roof and ceiling with space between them, air conditioners, fans etc)
2. There are windows that can be opened or there are air vents
3. Direct sunlight cannot enter the area (e.g. window panes are painted or there are curtains/blinds to protect against the sun)
4. Area is free of moisture (e.g. leaking ceiling, roof, drains, taps etc)
5. There is a cold storage in the facility
6. There is a regularly filled temperature chart for the cold storage
7. Medicines are not stored directly on the floor
8. Medicines are stored in a systematic way (e.g. alphabetical, pharmacological etc)
9. Medicines are stored first-expiry-first-out (FEFO)
10. There is no evidence of pest in the area
11. Tablet/capsules are not manipulated by naked hand
12. There are security measures to avoid burglary
13. Ledgers are up to date and complete
14. Storage equipment is adequate (shelves, pellets, etc)
15. Adequate storage space is available for drugs and medical supplies
16. Dispensing bags for medicines are available

807 *Appendix A.5. Results for tests on time coefficient with alternative inference strategies*

Variable	Test	DoF	t Statistic	P-value	P-value Inflation factor
Antibiotics	Alternative t test	6	3.55	0.00 ***	0.34
Antibiotics	Wild cluster	6	3.55	0.03 *	2.32
Antibiotics	Subcluster	225	2.40	0.002 **	1.66
Antimalarials	Alternative t test	6	2.63	0.33	8.63
Antimalarials	Wild cluster	6	2.63	0.10	2.34
Antimalarials	Subcluster	225	2.15	0.04 *	0.98
Other drugs/vaccines	Alternative t test	6	7.45	0.02 *	83.52
Other drugs/vaccines	Wild cluster	6	7.45	0.02 *	79.90
Other drugs/vaccines	Subcluster	225	4.56	0.00 ***	0.00
Reproductive health	Alt t test	6	2.17	0.04 *	0.54
Reproductive health	Wild cluster	6	2.17	0.14	1.89
Reproductive health	Subcluster	225	1.199	0.24	3.34
All drugs	Alternative t test	6	8.35	0.01 *	64.50
All drugs	Wild cluster	6	8.35	0.03 *	175.40
All drugs	Subcluster	225	3.81	0.00 ***	0.00

Cluster robust standard errors in parentheses; * < 0.05, ** < 0.01, *** < 0.001

Notes: (1) lines represent significance tests for time coefficients ("Post") estimated in Table 5, for the different drug categories considered in our analysis (antibiotics, antimalarials, other drugs and vaccines, drugs for reproductive health, all drugs pooled); (2) Three approaches to inference test are: t distribution with $G - 1$ degrees of freedom (where G is the number of clusters), wild clustered bootstrap and sub-cluster wild bootstrap. (3) The empirical approach is described in detail in section 3.