Health Economics WILEY

Impact of social accountability monitoring on health facility performance: Evidence from Tanzania

Igor Francetic[1,2,3,4] | Günther Fink[2,3] | Fabrizio Tediosi[2,3]

[1]Health Organization, Policy and Economics (HOPE) Group, Centre for Primary Care and Health Services Research, School of Health Sciences, University of Manchester, Manchester, UK

[2]Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland

[3]University of Basel, Basel, Switzerland

[4]Department of Business Economics, Health and Social Care, University of Applied Sciences and Arts of Southern Switzerland, Manno, Switzerland

Correspondence
Igor Francetic, Health Organization, Policy and Economics (HOPE) Group, Centre for Primary Care and Health Services Research, School of Health Sciences, University of Manchester, Oxford Road, Manchester, M139PL, UK.
Email: igor.francetic@manchester.ac.uk

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 160373

Abstract

Social accountability programs are increasingly used to improve the performance of public service providers in low-income settings. Despite their growing popularity, evidence on the effectiveness of social accountability programs remains mixed. In this manuscript, we assess the impact of a social accountability intervention on health facility management exploring quasiexperimental variation in program exposure in Tanzania. We find that the social accountability intervention resulted in a 1.8 SD reduction in drug stockouts relative to the control group, but did not improve facility infrastructure maintenance. The results of this study suggest that social accountability programs may be effective in areas of health service provision that are responsive to changes in provider behavior but may not work in settings where improvements in outcomes are conditional on larger health systems features.

KEYWORDS
community monitoring, essential medicines, health infrastructure, social accountability, Tanzania

1 | INTRODUCTION

Social accountability programs are an increasingly popular tool for improving the performance of health systems in low- and middle-income countries (LMICs). By allowing patients and citizens to directly provide feedback to health workers, social accountability essentially empowers communities to exert social pressure on responsible health workers, increasing their incentives to meet local expectations. Social accountability programs come in many different forms in practice, and range 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 & Houtzager, 2012; Molina, Carella, Pacheco, Cruces, & Gasparini, 2017; Ogbuabor & Onwujekwe, 2018). Several earlier studies have linked accountability mechanisms to a wide array of positive outcomes at the health systems level, including increased service accessibility, better quality of care, increased patient satisfaction and provider performance, lower corruption, and ultimately better health outcomes (Greer, Matthias, Josep, & European Observatory on Health Systems and Policies, 2016; Kalolo, Radermacher, Stoermer, Meshack, & De Allegri, 2015; Lewis, 2006; Lewis & Pettersson, 2009). Fueled by successful programs like Uganda's community-based monitoring (Björkman & Svensson, 2009), social accountability projects have been launched in many low-income healthcare settings in recent years (Molina et al., 2017).

Yet, the evidence on the effectiveness of social accountability programs remains mixed. Some further positive evidence on reduced corruption and increased service quality in Uganda (Fiala & Premand, 2018) as well as on improved access to maternal health services in India (Hamal, Tjard de Cock, De Brouwere, Bardají, & Dieleman, 2018) is contrasted by several studies finding no impact associated with social accountability initiatives (Gaventa & McGee, 2013; Lodenstein et al. 2017, 2018), including an attempt to replicate the original "power to the people" project results in Uganda (Björkman Nyqvist, de Walque, and Svensson 2017).

In this study, we assess the impact of a social accountability monitoring (SAM) program recently implemented in selected areas of Tanzania. Starting in 2011, a large *Health Promotion and System Strengthening* (HPSS) *Project* was launched in all seven districts of Dodoma region, Tanzania. Two out of seven districts in the region were selected for an additional SAM program. These SAM districts were similar to the five other (control) districts in the region in terms of demographic and socioeconomic characteristics, as well as with respect to their health systems characteristics at the beginning of the project. The main objectives of the SAM program implemented from 2012 were to increase the availability of medicines, to improve health facilities infrastructure maintenance, and to improve resource allocation as well as financial management at the district level. Using a difference-in-differences (DID) approach to identify the changes in the two SAM districts relative to the five districts not receiving the program over the periods 2011 and 2017, we find that the SAM intervention on average reduced the number of drug stockout days, but did not change facility infrastructure maintenance efforts. Our results suggest that stockouts across all drugs considered were reduced by 266 days over a 3-month period, a 1.8 standard deviation reduction relative to the stockout levels observed at baseline. Most of these reductions seem to be linked to stockouts in antibiotics and other essential drugs. We implemented an extensive series of robustness checks to address potential threats to our primary identification strategy. First, we address potential selection into treatment based on initial differences in study outcomes using a lagged dependent variable (LDV) approach. Second, we use drug availability from national surveys to estimate alternative models with matched control districts outside of the region. Thirdly, we estimate our main model selectively excluding treatment and control districts. Finally, we run placebo regressions, analyzing a set of alternative outcomes which should not have been affected by the policy. All of these tests substantially confirm our main results.

In the last part of the paper, we discuss potential mechanisms that may explain the incremental effect of the SAM project. Building also on qualitative data collected across a purposely sampled group of facilities in Dodoma region, we argue that the most plausible mechanism for the improved drug stocking outcomes seems to be a perceived increase in pressure on providers, who reacted to this increase in social pressure by intensifying efforts in domains under their direct influence, such as forecasting needs of medicines and filing timely and complete drugs orders. Other domains of health facility performance—where change potentially requires coordinated efforts with higher level authorities, substantial funding and planning—were likely harder to control for local staff and thus less likely to respond to social accountability.

The results presented in this study are consistent with the idea that increased transparency and social accountability can improve outcomes in settings where marginal changes in provider efforts can affect observable outcomes (Fiala & Premand, 2018; Fox, 2007, 2015; Joshi, 2017; Joshi & Houtzager, 2012; O'Meally, 2013). The main enabling factor highlighted in the literature that is supported by our analysis is the ability to accommodate context specific factors into the program (Danhoundo, Nasiri, and Wiktorowicz 2018; Gaventa & McGee, 2013; Martin Hilber et al., 2016). In the specific case analyzed here, the non-governmental Organization (NGO, Sikika) was able to embed the accountability process into existing social and institutional mechanisms. Several authors have also highlighted the importance of feedback cycles between community and government authorities (Fox, 2015; Ringold, Holla, Koziol, & Srinivasan, 2012).

The remainder of the paper is structured as follows: Section 2 introduces the study setting and the social accountability intervention, Section 3 discusses data and methods. We present our main results in Section 4 and our discussion in Section 5.

2 | STUDY SETTING AND SOCIAL ACCOUNTABILITY INTERVENTION

2.1 | Study setting

Our study was conducted in Tanzania. The country has an estimated population of 56 million and is administratively organized in 31 regions, further divided into about 170 districts councils. The Tanzanian health system is highly decentralized with many health system responsibilities delegated to local government authorities and health districts (Kigume and Maluka 2018b). Regional authorities maintain an overseeing role over districts and act as intermediaries with respect to the central government, which assigns the core budget based on broad population characteristics (Kigume and Maluka 2018b). Health care is organized with a pyramidal structure, with a large number of dispensaries (about 80% of all health facilities) offering primary care, local health centers offering secondary care (about 15%) and supervision to dispensaries, and individual (public, private or faith-based) district hospitals (about 4%) providing highest level care in each district (Musau et al., 2011).

The SAM program analyzed was implemented in the region of Dodoma. 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. The 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 *HPSS* program was implemented across all districts of Dodoma 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.[1]

Stockouts of essential medicines have been a central problem in Tanzania for many years. Appendix A1 offers an overview over medical supplies in Tanzania as well as specific national policies implemented to increase access to essential medicines in the last decade. In short, when drugs are not available at point of care, households typically buy them in drug stores, pharmacies or accredited dispensing outlets. These shops are abundant in urban areas but less accessible in most rural areas, with little difference across our study sites (Wiedenmayer et al., 2019). Larger pharmacies and wholesale drug stores are also rare across rural districts, which complicates the procurement process for health facilities experiencing shortages, as we discuss in length in Appendix A1. To address the problem, the HPSS Medicine Management component had a specific sub-component addressing health workers′ training, accountability, rational use of medicines and notably the medicines supply-chain (Wiedenmayer et al., 2019). This latter element was explicitly meant to address stockouts, inefficiencies and shortages in medicines deliveries attributable to the centralized government central Medical Store Department (MSD). The medical supply chain tool was a complementary supply of medicines through an innovative public-private partnership initiative known as "Jazia Prime Vendor" (PVS) system (Wiedenmayer, 2017). Under the PVS, district authorities were allowed to purchase medicines from a single private provider when the centralized MSD experienced shortages; in these cases, district authorities consolidate the orders placed by health facilities for the Prime Vendor, which then executes them. The single private supplier—the Prime Vendor—had to be tendered and contracted for the whole region based on price and quality attributes. District authorities were then responsible for the distribution of supplies purchased through the PVS to facilities. This system replaced the previous uncoordinated and inefficient procurement from local private providers to complement unmet needs related to MSD central stockouts. The Jazia PVS did not take off until the end of year 2014, with a pilot project implemented from September 2014 to July 2018 in all districts in the Dodoma Region simultaneously.

Besides the availability of medicines, HPSS also addressed maintenance of health facility infrastructure. The state of infrastructure and equipment in public health facilities across LMICs—together with the availability of trained staff, drugs, and health commodities—is a crucial prerequisite for the provision of quality healthcare (Penfold et al., 2013; van Pelt et al., 2020). Despite the importance of infrastructure and equipment, and in spite of large investments to extend geographical access to care, the level of investments in maintenance across Tanzanian health facilities remains inadequate (Scholz, Ngoli, & Flessa, 2015) especially for lower tier facilities (Boex, Fuller, & Malik, 2015).

No regional plans or guidelines for the maintenance of health technology and infrastructure were available in Dodoma region as of 2012. Planning and execution of infrastructure maintenance was generally delegated to health facilities, either with own funds or with funds obtained by district authorities through specific applications (Stoermer, Werlein, & Molesworth, 2011). The lack of personnel trained in infrastructure maintenance—as well as the high burden of care-related activities—severely impaired the ability of facilities to conduct proper maintenance. Starting from 2013, the HPSS program issued standard operating procedures, conducted inventories, gradually trained personnel on maintenance or hired specialized staff such as technicians and engineers (Stoermer et al., 2011).
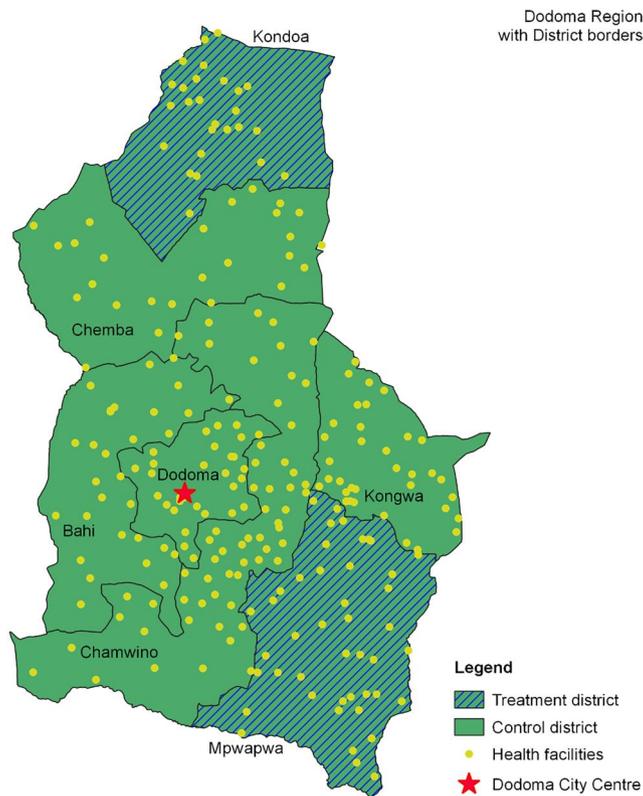
FIGURE 1 Districts in Dodoma Region and location of surveyed health facilities. Geographic Information System (GIS) layers with district boundaries are NBS (2017). GIS position for surveyed health facilities was recorded by the Health Promotion and System Strengthening project

2.2 | The intervention: SAM

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"[2]. With funding from several development partners, Sikika launched a SAM program targeting health service delivery in 2012. Sikika's SAM project was centered on an articulated accountability process that involved several 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. Based on available funding and existing links to these districts, Sikika decided to implement the SAM program in two out of the seven districts in Dodoma: Kondoa and Mpwapwa. As illustrated in Figure 1, the two districts lie at the Northern and Southern ends of the region, with the five other districts located in between these two areas providing some obvious reference or benchmark.

The outcomes explicitly targeted by the SAM program in the Dodoma region were the following:

- Reduction of stockouts 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 (e.g., Health Facility Governing Committees [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, that is, 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 multichannel media strategy aimed at facilitating the achievement of the above objectives (Sikika 2013).

Before starting the implementation of the SAM process in a new district, Sikika met with local government authorities and other stakeholders to introduce the SAM process and principles of the project. After this preliminary stage, community meetings were held to select democratically citizens that would join the district SAM team; each district SAM team was composed of 15–20 members, including citizens, representatives from district authorities (Council Health Management Team, Council Management Team), HFGCs, religious groups, local NGOs or other grass-roots level CSOs. Each SAM team was responsible for the implementation of the SAM process, covering all government managed health facilities within the district boundaries (in our case 20–50 depending on the district). Once the formation of SAM teams was complete, members followed 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. Full details are reported in Appendix A2.

As illustrated in Table 1, the SAM program was launched in 2012, simultaneously to the launch of HPSS project activities (that were implemented in all seven districts). Up to September 2014, the HPSS program implemented only inventory and training activities related to the health technology and infrastructure component. From September 2014 to May 2017, Sikika's SAM also overlapped with the HPSS component addressing the medicines supply-chains, namely the Jazia PVS. Given that these HPSS programs were run in all districts of Dodoma, our estimated impact of the SAM program should be interpreted as incremental benefits of the SAM intervention in the context of larger health system reform efforts.

Sikika's SAM initiative was supposed to yield increased health workers' effort in forecasting needs and filing timely orders at the level of health facilities. At the level of districts, the SAM program was meant to help highlighting and demanding improvements on the failing elements in the procurement and delivery chain, as well as to increase efforts to promote facility maintenance and investment through mobilization of district level funds.

Table 2 compares control and treatment areas at baseline with respect to population, health system and health facilities characteristics. Overall, treatment and control areas seem very similar at baseline—none of observed differences at baseline appear to be statistically significant. The only variable where substantial (even if not statistically significant) differences were observed for health insurance enrollment. The higher baseline health insurance coverage in the treatment group appears to be driven by Kondoa district, where district authorities actively promoted enrollment in the community health insurance scheme for all individuals covered by the national poverty alleviation scheme (Tanzania Social Action Fund [TASAF]; Sikika 2016b). Higher insurance coverage is also reflected in a higher amount of funds disbursed by the health insurance schemes across treatment districts, which could in turn improve local availability of funds and infrastructure maintenance. As described in further detail in Section 3.2, we control for per capita CHF funds disbursed to facilities at the district level in our analysis to reduce potential confounding or omitted variable concerns.

3 | DATA AND METHODS

3.1 | Data

The main data sources for our analysis are two health facility surveys conducted as part of the HPSS project activities.

HPSS surveyed randomly selected households and health facilities in Dodoma region 2011 (baseline) and in 2017 (endline). Both health facilities surveys covered only government-managed health facilities; private and faith-based facilities were neither included in the HPSS project nor targeted by the SAM program. As of 2012, the Dodoma region had 267 government-managed health facilities. The number rose to about 286 in 2017 (Ministry of Health, Community Development, Gender, Elderly and Children 2019; Ministry of Health and Social Welfare 2011). The 2011 baseline HPSS survey included all government-managed health facilities in the region. For the 2017 survey, HPSS sampled randomly about half of the health facilities in the region, stratifying by district and health facility type (Kuwawenaruwa, Wyss, Wiedenmayer, Metta, & Tediosi, 2020; University Consultancy Bureau 2018). The full samples resulting from preliminary data cleaning procedures include 112 health facilities at endline, 91 of which were observed at both time points forming the balanced panel on which we focus in our main analysis. We did not detect systematic characteristics in the facilities excluded at data cleaning stage which could generate bias in the results, including the facilities lost due to attrition. Table 3 shows the composition of our main sample, while Appendix A3 reports specific compositions of the two cross-sectional samples (baseline and endline).

TABLE 1 Timeline of interventions in Dodoma region

Year	Month	Actor	Event
2011	Aug–Sep	Health Promotion and System Strengthening (HPSS)	Baseline survey in Dodoma region
2012	Oct	Sikika	First SAM round starts in Kondoa and Mpwapwa districts
2013	Jan	HPSS	First training and inventory activities for health technology and infrastructure
2014	Feb	Sikika	Second social accountability monitoring (SAM) round starts in Kondoa and Mpwapwa districts
2014	Sep	HPSS	Jazia Prime vendor (PV) system starts operations to address stock outs
2015	Jan–Dec	Sikika	Follow up and monitoring activities in Mpwapwa
2016	Jan–Dec	Sikika	Follow up and monitoring activities in Mpwapwa
2016	Aug	Sikika	Third SAM round starts in Kondoa
2017	Jan–Jun	Sikika	Follow up and monitoring activities in Kondoa and Mpwapwa
2017	May–Jun	HPSS	Endline survey in Dodoma region

Sources: Stoermer et al. (2011); Sikika (2013, 2014, 2015, 2016a); Wiedenmayer (2017).

Our first outcome of interest was drug stockout days. The HPSS survey collected detailed data on a number of essential medicines (University Consultancy Bureau 2018). After data cleaning and subsequent cross-validation with HPSS staff, we included 13 medicines in our analysis and computed total stockout days as the (cumulative) sum of days these medicines were out of stock (duration) during the 3 months prior to the survey. The 90 days reference period should offer a good proxy indicator of average stocking behavior at the surveyed facilities, and can also be viewed as a measure of the effective treatment days lost as a result of lack of appropriate medicines. Appendix A4 shows the main drug groups included in our analysis, which include four antibiotics, two antimalarials, four drugs specifically used in reproductive health and three other essential drugs and vaccines.

Our second outcome of interest was infrastructure maintenance. To assess infrastructure, we extracted data on drug storage infrastructure availability from a wider check-list of infrastructure maintenance practices. These questions were asked by independent surveyors to health facility staff as part of the HPSS health facility surveys. Among the 16 questions listed in Appendix A4, we focused on the items that seemed most closely linked to the program's goals, that is, (1) functioning equipment to maintain cold storage for vaccines and medicines, and (2) adequate furniture and equipment to store properly medicines. The SAM program also targeted financial management and governance. These outcomes were however not as clearly defined and thus not included in our analysis.

We complemented the HPSS survey data with multiple additional sources, listed in Appendix A6. We included several contextual factors—at health facility 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, location in urban versus rural area, distance to the city center of Dodoma as proxy for the distance from the central zonal MSD store and average yearly rainfall (Wagenaar et al., 2014). In this context, the inclusion of average rainfall as covariate is motivated by three factors: (1) reduced ease of access to health facilities for households during rainy season may exogenously curb demand, (2) traveling during rainy season in areas with few tarmac roads can affect negatively the ability of MSD to deliver drugs, and (3) areas with higher yearly rainfall are characterized by higher malaria prevalence, which exogenously increases demand of antimalarials and other generic drugs to treat children (Adhvaryu & Nyshadham, 2015; Moïsi et al., 2010). To control for the pace of implementation of the HPSS program and the possible influence on our outcome variable, in our analysis we also added a dummy variable indicating facilities reporting at least one HPSS supportive supervision on medicine supply in the previous quarter. At the district level, we used Annual Health Statistics developed by the Tanzanian Ministry of Health and Social Welfare (MoHSW) and National Bureau of Statistics (NBS) to obtain data on population, density of health facilities and outpatient visits. To account for household trends in wealth, we also constructed district wealth index quintiles using data made available by the local NGO Twaweza. Since 2009, with support of the British, Danish and Swedish development agencies, Twaweza implemented an educational program (Uwezo) and contextually collected district-representative households surveys from 2009 to 2017 (Sumra, Ruto, and Rajani 2015). Finally, in the Tanzanian decentralized health system, both the availability of essential medicines and infrastructure maintenance depend—with varying degrees—upon the availability of locally generated

TABLE 2 Differences between control and treatment group at baseline (2011)

Variable[a]	Control group		Treatment group		Difference[c]	
	Obs.	Avg./Perc.	Obs.	Avg./Perc.	(C–T)	*p*-value
Delivery at health facility (perc.)	217	37.79	91	36.26	1.53	0.872
Children birth weight (g)	94	3215.43	37	3213.51	1.91	0.987
Children received Tuberculosis vaccine (perc.)	206	94.17	89	93.26	−0.91	0.793
Children received first vaccine for diphtheria, pertussis and tetanus (perc.)	206	92.23	89	93.26	−1.02	0.483
Children received first polio vaccine (perc.)	206	90.78	89	93.26	−2.48	0.267
Children received measles vaccine (perc.)	205	71.22	88	67.05	−4.17	0.400
District population (number)	147	311704	58	301258	10,446	0.727
District pop. in poorest wealth quintile (perc.)	147	24.88	58	31.33	−6.46	0.400
District pop. in richest wealth quintile (perc.)	147	15.24	58	11.07	4.17	0.466
District health insurance coverage (perc.)	147	0.13	58	0.27	−0.14	0.308
Districts funds disbursed by Community Health Fund per capita (TZS)[b]	147	158.78	58	209.39	−50.62	0.592
District health facility density (per 100′000)	147	16.07	58	15.70	0.36	0.779
District outpatient department (OPD) visits per 100′000 pop.	147	4389.03	58	4527.37	−138.34	0.868
Active health facility committee (perc.)	147	0.92	58	0.88	0.04	0.707
Urban area (perc.)	147	0.03	58	0.05	−0.02	0.233
Average yearly rainfall (mm per year)	147	560.36	58	630.88	−70.53	0.185
Share of exempted patients (perc.)	141	3.35	58	3.05	0.30	0.422
Facilities quarterly supervised on medicine supply (perc.)	146	0.66	58	0.71	−0.05	0.844
Facilities with staff trained on maint. (perc.)	141	0.55	58	0.64	−0.09	0.534
Antibiotics (days out of stock)	147	62.37	58	91.10	−28.74	0.387
Antimalarials (days out of stock)	147	51.32	58	47.38	3.94	0.701
Drugs used in reproductive health (days out of stock)	147	44.33	58	58.19	−13.86	0.133
Other essential drugs and vaccines (days out of stock)	147	34.82	58	43.15	−8.33	0.378
All drugs (days out of stock)	147	192.84	58	239.83	−46.99	0.263
Cold storage (e.g., fridge) is available	147	0.62	58	0.81	−0.19	0.086
Adequate furniture and equipment is available	147	0.39	58	0.26	0.13	0.517

[a]Detailed data sources are listed in Appendix A6.

[b]The exchange rate as of end of 2018 was 1 USD = 2298 Tanzanian Shillings (TZS).

[c]Slight inaccuracies in the reported differences (C–T) are due to rounding. Test for differences in means or proportions with robust errors clustered by district. We reported the *p*-values in the last column.

funds (Kuwawenaruwa et al., 2020). In the case of Dodoma, these result either from user fees or from funds collected and disbursed by the local community health insurance scheme, known as the CHF. Heterogeneity in local availability of these funds might cast doubts about our identification strategy. Therefore, we obtained data on the amount of funds (per capita) disbursed by the CHF to health facilities across the district.

3.2 | Empirical approach

Our study aims at assessing the impact of SAM on health facility performance, measured as duration of stockout for 13 essential medicines and two indexes representing maintenance of infrastructure (drug storage and dispensing areas). Our main DID approach explores differences between treatment and control areas over time. Our treatment group

	Dispensary N (%)	Health center N (%)	District hospital N (%)	Total N (%)
Baseline (pre)	56 (84.9)	8 (12.1)	2 (3.0)	66 (100.0)
Control				
Treatment	20 (80.0)	11 (12.0)	2 (8.0)	25 (100.0)
Endline (post)	57[a] (86.4)	7[a] (10.6)	2 (3.0)	66 (100.0)
Control				
Treatment	20 (80.0)	11 (12.0)	2 (8.0)	25 (100.0)

TABLE 3 Composition of main sample (balanced panel) at baseline and endline

[a]One facility classified as health center at baseline was re-classified as (downgraded to) dispensary at endline—we treated this facility the same in both survey rounds.

includes 25 facilities in the districts of Kondoa and Mpwapwa, where the Sikika program was rolled out. In our main specification, the control group includes 66 facilities located in the other five districts in Dodoma region (Figure 1). The full model is given by:

$$Y_{i,d,t} = \alpha_i + \beta_0 + \beta_1 \, Post_t + \beta_2 \, Treatment_d + \delta \, Post_t \times Treatment_d + \beta_3 \, X_{i,t} + \beta_4 \, \Omega_{d,t} + \epsilon_{i,d,t}$$

where, $Y_{i,d,t}$ is the outcome of interest for health facility i in district $d \in D$ for period $t \in (2011, 2017)$. $Post$ is a binary indicator for the time period (either 0 for the pretreatment 2011 baseline or 1 for the post-treatment 2017 endline). $Treatment$ is an indicator variable that equals 1 for health facilities in the treatment districts and zero for those included in the control group. $X_{i,t}$ is a matrix of observable health facility control variables, while $\Omega_{d,t}$ is a matrix of district level control variables. β_3 and β_4 are the vectors of coefficients for the control variables and α_i is a vector of time-invariant health facility fixed characteristics (facility-specific intercepts). Finally, $\epsilon_{i,d,t}$ is the idiosyncratic error. Our coefficient of interest—capturing the additional effect of the SAM program—is represented by δ.

To ensure that our results are not affected by differences at the health facility level, we exploit the longitudinal dimension of our data and estimate our main model using a fixed effects (FE) within-estimator, averaging out any time-invariant observed and unobserved facility characteristics to remove this potential source of bias. The primary identifying assumption in our main model is thus of independence of the SAM treatment conditional on unobserved time-invariant facility effect as well as on time-variant observable facility and district characteristics.

From an inference perspective, the small number of groups (seven districts) in our analysis poses an additional challenge (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 the permutation-based wild clustered bootstrap (Cameron, Gelbach, & Miller, 2008) with a 6-point weight distribution (Rokicki, Cohen, Fink, Salomon, & Beth Landrum, 2018; Webb, 2013) and the sub-cluster wild bootstrapping approach recommended by (MacKinnon & Webb, 2018).

The overlap between HPSS and Sikika's activity casts doubts about the potential confounding effect of HPSS on our results. To address this issue, we report our analyses with and without process variables capturing the extent of HPSS activities. If these activities were the true drivers of changes observed, then estimated treatment effects should shrink or disappear when these efforts are controlled for empirically. All data management and analysis was conducted using the statistical package Stata, version 14.

3.3 | Addressing threats to causal inference

The proposed identification strategy in our primary DID rests on the key assumption of parallel trends in treatment and control groups in absence of the treatment (Dimick & Ryan, 2014), which is hard to verify even if pretrend data was available. Differences in trends may arise due to different long-term trajectories, but also due to differential governmental or NGO efforts in treatment or control areas. Given that we cannot formally test the common trends assumption, we test whether our main results are robust to alternative model specifications that do not rely on common trends as central assumption (Angrist & Pischke, 2009). In light of these limitations—although we rely on the FE DID as

main estimator due to its intuitive interpretation in our setting—the results should be interpreted jointly with the following alternative specifications.

3.3.1 | Controlling for pretreatment differences with a LDV approach

Although the DID identification strategy technically does not require baseline balance between outcomes in treatment and control groups, the initial differences in medicines availability highlighted in Table 2 raise concerns regarding potential convergence effects. Even in the absence of additional interventions it seems possible that poorly performing health facilities in treatment areas could have experienced faster improvements compared to facilities in control areas due to convergence or mean reversion patterns in the data. To address this concern, we estimate LDV (or analysis of covariance) models that explicitly control for the initial levels of stockout days and only assume independence conditional on past outcomes and covariates (O'Neill, Kreif, Grieve, Sutton, Jasjeet, & Sekhon, 2016). Besides having less restrictive identifying assumptions and controlling for potential baseline differences, this approach can achieve greater power compared to DID, which is interesting provided our limited sample size (Raza, Van de Poel, and Van Ourti 2018). Additionally, FE and LDV estimates can be used for bracketing likely ranges, providing plausible lower (LDV) and upper (FE) bounds to the casual effects of interest (Angrist & Pischke, 2009; Ding and Li 2019).

3.3.2 | Matching control districts outside Dodoma based on past outcomes and covariates

Besides imbalances in baseline characteristics, which we try to address with the LDV approach, it seems possible that the two districts chosen for the treatment were somewhat special with respect to their basic infrastructure and supply chains. We thus pursue an alternative matched control approach to support our main results, broadly following previous work by O'Neill at al. (2016). Both the 2012 Service Availability and Readiness Assessment (SARA) and the 2014 Service Provision Assessment survey (SPA) collected nationally representative data on the availability of essential medicines on the day of the survey, including the 13 drugs described in Appendix A4. Using these data sets, we constructed availability indicators (i.e. the number of medicines available on the survey date, out of the 13 considered) for all health facilities included in our original data as well as for additional facilities in other regions and districts across Tanzania. The resulting dataset features repeated cross-sections from several surveys which we could not link longitudinally: 2011 HPSS baseline survey, 2011/2012 SARA, 2014 SPA and 2017 HPSS endline survey. The primary drug availability index is different from the one used in the main model because it only covers drug availability on the day of the interview (rather than availability over 90 days). Conceptually, this measure may be less prone to recall biases, but is more likely to be affected by random within-period variation in stocking, and thus likely does not capture average stock availability as well as the measure available in the HPSS survey data.

In order to make this approach fully complementary to our main approach (and to rule out local spillover effects), we excluded original control districts, as well as districts where Sikika implemented similar SAM activities after 2012 as well as two regions with other HPSS program (Morogoro and Shinyanga) from this matching exercise. This allows additionally to assess the program impact over a time span when the main HPSS medicine management component was not in place yet. We then computed district level Malahanobis distance between potential controls and the two treatment districts at baseline (Kantor, 2012). Multidimensional distance was computed over a vector of baseline district characteristics, including: the outcome of interest, share of facilities within the district located in urban areas, district population, district malaria prevalence, as well as share of population in lowest and highest wealth index quintiles. Further details are reported in Appendix A7. Based on the resulting Mahalanobis scores, we selected the 10 most similar control districts. Figure 2 shows the location of our original treatment and new matched control districts. Baseline average availability for our set of essential medicines was 9.20 (st. dev. 0.24) for control and 8.74 (st. dev. 0.49) for treatment districts.

3.3.3 | Alternative treatment and control groups

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 for the main pooled drug stockout days outcome. It seems possible, for example that the intervention districts in the South (Mpwapwa) or North (Kondoa) could have
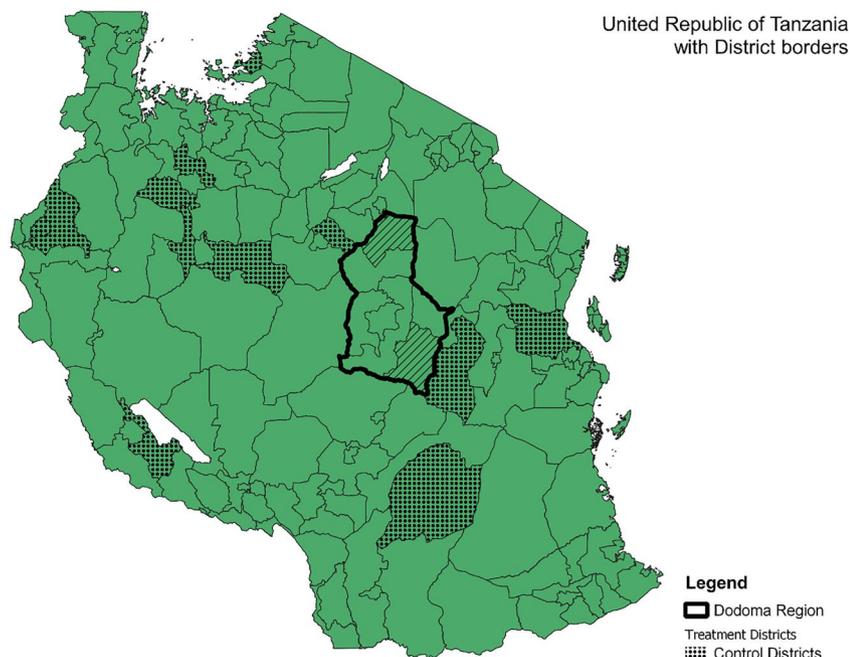
FIGURE 2 Location of treated and matched control districts. GIS layers with district boundaries are National Bureau of Statistics (2017)

differentially been affected by health insurance coverage (see Table 2) or that control areas in City of Dodoma—as capital city of the United Republic of Tanzania—have been affected by other governmental or political efforts.

3.3.4 | Additional approaches to assess the reliability of our main specification

To further attenuate concerns regarding other unobserved factors affecting the treatment districts, we propose a falsification test based on placebo regressions (Athey & Imbens, 2017; Egedesø, Hansen, and Jensen 2020). In absence of pre- or post-treatment data which allow directly testing for differential trends, we analyze two variables that should not show any response to the SAM program over the study period unless there were other policy initiatives or local changes affecting the health system. Finding an intervention effect on these placebo outcomes would thus suggest other programs or factors driving differences between districts, other than the changes introduced by the SAM program. First, we look at the share of people exempted from user fees at the surveyed facilities. This variable captures both local poverty levels and general health system financing efforts by the government that could create differential demand across districts, and potentially explain our main results. Second, we look at the share of households reporting children under five sleeping under insecticide-treated bed nets, which we consider both a proxy for (vertical) program-specific efforts by the government, but could of course also predict the demand for antimalarials (where we see large effects) at facilities.

One may still be concerned that the observed impact on drug availability may be related to a differential efficiency of the governmental medicine supplier (MSD) in the first place. While higher MSD order fulfillment would certainly decrease stockouts, the centralized structure of MSD makes the local SAM very unlikely to affect this indicator. We therefore looked at the treatment effect on a dummy variable indicating whether the last MSD order (previous to the survey) was completely fulfilled across the facilities included in our main analysis.

As mentioned above, we do not have data on facility stocking or infrastructure prior to 2012 that would allow us to directly analyze pretrends. In Appendix A13 we include an attempt to assess parallel pretrends using two indicators that should reflect health system efforts as well as general disease ecology.

3.4 | Simulated power calculation

Given that our analysis is based on data which was not collected for the primary purpose of evaluating the impact of Sikika's SAM program, ex ante power calculations are not available. To provide some sense of required effect

sizes with the data set at hand and error corrections applied, we simulate power for our main outcomes in Appendix A8. Starting from the original baseline values for our core sample of health facilities, we simulate endline values using an AR (1) data generating process calibrated to approximate endline control group average values. We then empirically assessed the likelihood of obtaining p-values below the 0.05 threshold for a range of plausible effect sizes. The results of these simulations suggest that our main FE specification with 91 facilities should have allowed us to detect effect sizes of 0.5 standard deviations with probability 0.8, and effect sizes of 0.65 standard deviations with power 0.9.

4 | RESULTS

4.1 | Main results

Table 4 presents the results of our main FE DID specification. We estimate that the Sikika SAM program reduced antibiotic stockout days by 118 days (out of a potential maximum of 360), reproductive health stockout days by 70 (potential maximum 360), and other essential drug stockout days by 76. Overall stockout days were reduced by 266 days over a 3-month period, which means approximately three additional medicines available on each of the 90 days considered. Relative to baseline, this corresponds to a 1.6 SD improvement in antibiotics availability, a 1.1 SD improvement in reproductive health supplies and a 1.8 SD improvement in the availability of other essential drugs and for overall drugs. The inclusion of HPSS implementation controls in the bottom panel of Table 4 affected the estimated effects only marginally, suggesting that HPSS activities do not substantially interfere in the identification of the SAM impact.

Table 4 also reports bootstrapped p-values based on the wild cluster bootstrap procedure in addition to the cluster-robust standard errors (CRSEs) used in our main model. In Appendix A10 we report p-values under a range of alternative approaches that have been proposed in the literature for settings like ours with a small number of clusters: a t distribution with six degrees of freedom, wild clustered bootstrap and wild bootstrap at sub-cluster level. Under CRSEs, estimated impacts are statistically significant at $p < 0.05$ for pooled stockout days, antibiotics, drugs used for reproductive health as well as other essential drugs or vaccines. Without HPSS controls, and using a cutoff value of 5%, the policy effect is significant under alternative t test approach for all drug groups except antimalarials. The p-values under wild cluster bootstrap were considerably larger than standard CRSE, yielding p-values < 0.05 only for antibiotics and all drugs pooled. Overall, significance for subgroups of medicines varied quite a bit with the approach chosen, while the estimated impact on total stockout days, antibiotics and other essential drugs or vaccines were so large in magnitude that they remained significant independent of the correction procedure applied.

In Table 4, we also included results of FE estimates for the indicators related to drug storage infrastructure availability. We do not find any impact of SAM on drug storage infrastructure relative to facilities in control districts, which suggests a stark difference in the impact of SAM across areas of health facility performance. The alternative coefficient testing procedures listed above also point to null results. We discuss these differences between the drug availability and infrastructure results in further detail below.

4.2 | Robustness checks

4.2.1 | LDV approach

Table 5 shows the regression results for our same sample of 91 facilities for pooled drug stockout days and the two infrastructure variables. Given that this analysis focuses on changes in the outcome as a function of initial levels, we have only one observation per facility. The LDV estimates suggest a 200-day decline in overall stockout days, which is marginally smaller than our primary DID estimate, but not statistically different in terms of magnitude. Similar to our main results, no impact was found for infrastructure indicators. For both classes of dependent variables, the lagged coefficient is not significant, suggesting that mean reversion only plays a limited role in this setting. The high degree of similarity between our FE and LDV estimates suggest that the true causal effect should be relatively close to the 1.8 standard deviation effect reported in our main specification.

TABLE 4 Change in drug stockout days and drug storage infrastructure 2011–2017 in core health facilities—main model with facility fixed effects

	(1) Antibiotics[a]	(2) Antimalarials[a]	(3) Other essential drugs and vaccines[a]	(4) Reproductive health[a]	(5) All drugs[a]	(6) Cold storage (e.g., fridge) is available[b]	(7) Adequate furniture and equipment is available[b]
Models without Health Promotion and System Strengthening (HPSS) control[c]							
Post X treatment[h]	−118.1*** (30.40)	−1.371 (29.34)	−76.31*** (15.74)	−70.36** (25.18)	−266.2** (80.25)	−0.393 (0.290)	−0.0813 (0.0620)
Bootstrapped p-value [d]	0.02**	0.97	0.02**	0.09*	0.05**	0.87	0.27
Adjusted R2	0.199	0.403	0.179	0.166	0.198	0.479	0.207
Akaike Information Criterion (AIC)	1863.3	1720.4	1708.9	1812.0	2126.9	−6.036	61.01
Bayesian Information Criterion (BIC)	1882.5	1739.6	1728.2	1831.2	2146.1	13.19	80.23
Observations[e]	182	182	182	182	182	182	182
Health facility controls[f]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District controls[f]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HPSS controls[c]	No	No	No	No	No	No	No
Models including HPSS control[b]							
Post X treatment[h]	−118.0** (33.14)	1.849 (25.49)	−73.87*** (18.45)	−60.79 (28.48)	−250.9** (88.54)	−0.384 (0.295)	−0.0360 (0.0650)
Bootstrapped p-value[d]	0.04**	0.96	0.04**	0.14	0.09	0.90	0.49
Adjusted R2	0.206	0.391	0.174	0.198	0.214	0.473	0.218
AIC	1831.6	1693.8	1680.8	1766.5	2086.9	−6.480	54.49
BIC	1850.7	1712.9	1699.9	1785.7	2106.0	12.64	73.61
Observations[e]	179	179	179	179	179	179	179
Health facility controls[f]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District controls[f]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HPSS controls[c]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline average[g]	75.15	53.77	32.65	53.47	215.04	0.68	0.36
Baseline standard dev.[g]	72.89	50.36	40.88	63.35	146.76	0.46	0.48

[a]Stockout days were computed over the 3 months (90 days) prior to the survey dates (September 2011 for baseline and May 2017 for endline). Composition of drug groups in Appendix A4.

[b]The variables represent binary responses to a check-list related to facility infrastructure availability, status and maintenance. The full check-list is available in Appendix A5.

[c]The HPSS implementation control variable is a dummy indicating facilities with supportive supervision for medicine management at least quarterly.

[d]Bootstrapped p-values referred to the wild cluster procedure.

[e]Sample includes all health facilities observed in both baseline (2011) and endline (2017) surveys.

[f]Control variables included: active health facility committee, average yearly rainfall, share of people in poorest wealth index quintiles, density of health facilities in district, number of OPD visits in district, district per capita Community Health Fund funds disbursed to health facilities.

[g]Baseline average and standard deviation referred to the dependent variable for the sample of facilities for year 2011.

[h]Estimates based on fixed effects (within) estimator. Full results including coefficients for covariates included in Appendix A9.

Cluster robust standard errors in parentheses: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

TABLE 5 Robustness check, lagged dependent variable estimation

	(1) Stockout days across all drugs[a]	(2) Cold storage (e.g., fridge) is available[b]	(3) Adequate furniture and equipment is available[b]
Post X treatment[g]	−200.3*** (50.17)	0.390* (0.172)	0.180 (0.238)
Bootstrapped p-value[c]	0.041**	0.36	0.68
Adjusted R^2	0.291	0.444	0.021
AIC	1064.5	20.64	131.2
BIC	1079.6	35.71	146.3
Observations[d]	91	91	91
Health facility controls[e]	Yes	Yes	Yes
District controls[e]	Yes	Yes	Yes
Health Promotion and System Strengthening controls	No	No	No
Baseline average[f]	215.04	0.68	0.36
Baseline standard dev.[f]	146.76	0.46	0.48

[a]Stockout days were computed over the 3 months (90 days) prior to the survey dates (September 2011 for baseline and May 2017 for endline) across all 13 drugs considered and listed in Appendix A4.

[b] The variables represent binary responses to a check-list related to facility infrastructure availability, status and maintenance. The full check-list is available in Appendix A5.

[c]Bootstrapped p-values are based on the wild cluster bootstrapping.

[d]Sample includes all health facilities observed in both baseline (2011) and endline (2017) surveys.

[e]Control variables included: active health facility committee, urban/rural area, facility type, distance to Medical Stores Department (MSD) warehouse, share of people in poorest wealth index quintiles, density of health facilities in district, number of OPD visits in district, district per capita CHF funds disbursed to health facilities.

[f]Baseline average and standard deviation referred to the dependent variable for the sample of facilities for year 2011.

[g]Estimates based on ordinary least squares (OLS) estimator. Full results including coefficients for covariates included in Appendix A12.

Cluster robust standard errors in parentheses: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

4.2.2 | Matched control approach

Table 6 shows the results of a DID model estimated using a multilevel specification with random intercepts (districts and regions) to account for the nested structure of the data. We ran the analysis over two time frames. Firstly, we restricted the study period to years 2011–2014, that is, the period before the alternative supply chain program in Dodoma (JPV) started. Secondly, we ran the same analysis for the full 2011–2017 period. The results are substantially aligned with our main analysis, suggesting increased availability of medicines in treatment districts. Interestingly, as anticipated, the effect appears higher for the analysis that includes the HPSS endline 2017 data. On average, we find that the SAM program increased the short-term availability of drugs by 1.5 drugs for the period 2011–2014, which corresponds to 0.82 standard deviations. For the full period 2011–2017 the effect increases to about 2 drugs, or 1.1 SD. While similar in relative magnitude, the effects estimated in Table 6 are not directly comparable to our main results in Table 4 because the SPA only measures availability on the day of the survey rather than stockout days over a 90 day period; both larger and smaller treatment effects over a longer reporting period definitely seem possible in principle.

4.2.3 | Alternative treatment and control groups

Appendix A12.4 reports full results for our analyses on drug stockouts with (1) treatment facilities only from Mpwapwa district, (2) treatment facilities only from Kondoa district, and (3) a control group excluding facilities in Dodoma Urban district. Using only Mpwapwa as treatment district we identify a 0.96 SD effect size; using only Kondoa as treatment district the effect is 1.15 SD. The estimated effect on the sample excluding the predominantly urban district of Dodoma 1.02 SD; none of the estimated coefficients is statistically different from our main estimates in Table 4.

	(1) Pre-Jazia PVS period 2011–2014[a,b]	(2) Full period 2011–2017[a,b]
Post X treatment group[f]	1.541[**] (0.607)	1.940[***] (0.594)
AIC	1679.8	1784.6
BIC	1712.1	1817.5
Observations[c]	421	450
Health facility controls[d]	Yes	Yes
District controls[d]	Yes	Yes
Baseline average[e]	9.13	9.13
Baseline standard dev.[e]	1.88	1.88

TABLE 6 Robustness check, changes in drug availability 2011–2014 and 2011–2017 with matched outside-of-region control district approach

[a]The dependent variable represents the number of drugs available on the day of the interview computed across the 13 drugs considered in Appendix A4, based on identical questions included in the different surveys. During the period 2011–2014, the Health Promotion and System Strengthening (HPSS) medicine management component (Jazia PVS) has not been rolled out. The full period 2011–2017 includes 3 years (2015, 2016 and 2017) with overlap of potentially beneficial effects from Jazia PVS and our intervention of interest (SAM).

[b]Matching of control districts outside of the Dodoma region based on smallest Mahalanobis distance computed on the following baseline district characteristics: outcome variable, share of facilities in urban areas, district population, malaria prevalence and share of people in lowest and highest wealth index quintile. The procedure is described in Appendix A7. The location of the 10 matched control districts is highlighted in Figure 2.

[c]Analysis based on pooled observations from HPSS (2011 and 2017), SARA 2012 and SPA 2014 surveys. See Appendix A6 for further details about data sources.

[d]Control variables included: facility type, health facility density, district population, urban/rural area, share of population in poorest and richest wealth index quintile.

[e]Baseline average and standard deviation referred to the dependent variable for the full sample of facilities for year 2011.

[f]Estimates based on multilevel model with random intercept (districts and regions), estimated using Maximum Likelihood with Expectation-Maximization algorithm. Complete results including coefficients for covariates and results of pooled OLS models with district fixed effects included in Appendix A12.

Heteroscedasticity-robust standard errors in parentheses; *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

4.2.4 | Additional robustness checks

The tables in Appendix A12 also report full results for our additional robustness checks introduced in Section 3.3.4. In our placebo regressions (A12.5) we did not detect any meaningful differences between treatment and control areas. Similarly, the results reported in A12.6 reveal no association between SAM and MSD order fulfilment.

5 | DISCUSSION

In this study, we evaluated the effect of a SAM program in two mostly rural districts of Tanzania. Overall, our results suggest that SAM had a positive and significant impact on drug availability—with particularly large effects on antibiotics and few other essential medicines. Interestingly, we found no program impact on antimalarials, which are in high demand in this area. One possible explanation for this is that malaria drugs are often managed outside government channels by large multilateral initiatives and vertical programs. All other categories of drugs—including antibiotics and other essential medicines/vaccines—are instead procured and distributed to government-managed health facilities through the regular supply-chain, which can potentially be influenced more easily by forecasting and order filing behavior of facility staff.

Our analysis did not find effects for other dimensions of health facility performance addressed by the SAM program. The analysis conducted on the indicators of drug storage infrastructure showed null effects with both the FE and LDV approaches. Unfortunately, we did not have reliable quantitative data for financial management, utilization of funds at

local level and functioning of health facility committees. We also did not have very detailed measures of infrastructure and efforts to maintain infrastructure—efforts in particular are rather hard to capture in routine facilities surveys. This clearly limits the scope of the measured impact of Sikika's program and the general conclusions to be drawn on the ability of the SAM program to improve health systems performance more broadly.

In addition to its limited ability to directly measure all four dimensions targeted, our study has some relevant limitations. Firstly, it is based only on two time points, that is, baseline and endline. The analysis would have benefited from the availability of preintervention 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. Secondly, our analysis represents a post-hoc analysis of a project implemented independently with nonexperimental treatment assignment and without a preanalysis plan. This limited our study in several ways, mainly restricting our ability to control for all possible confounders and producing analyses with inherently limited statistical power. 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 ways systematically correlated with the program assignments. Future research would benefit from an earlier involvement between researcher and implementation projects such as HPSS. This would allow the planning of a solid strategy to collect quantitative and qualitative data with relevant control variables and adequate pretreatment and intermediate outcomes (Boydell, McMullen, Cordero, Steyn, & James, 2019; Leatherdale, 2019). Lastly, similarly to other existing studies assessing impact of social accountability (e.g., McCoy, Hall, & Ridge, 2012; Molina et al., 2017; O'Meally, 2013), the results of our study, that is focused on a single region in Tanzania which is undergoing broader health reforms, may have somewhat limited external validity.

Despite these limitations, the results presented in this paper support the idea that SAM approaches might be effective in improving health systems performance fostering provider efforts. The magnitudes observed in our study are similar to those found in (Björkman & Svensson, 2009) as well as the results from a more recent pay for performance scheme in Tanzania (Binyaruka & Borghi, 2017).

One likely important contextual element is the positive and receptive administrative environment created by the region-wide implementation of the HPSS project, which likely facilitated systematic change in facility operations. Locally, Sikika is a well-known "watchdog" organization that was able to connect relatively easily with district authorities through the SAM process itself. The multiple rounds of discussion between the SAM teams and district authorities as well as the feedback to communities likely contributed to the positive outcomes for essential medicines as well as the generally positive attitudes toward the program seen in focus group discussions and interviews with facility staff. The SAM process introduced a structured bottom-up and top-down feedback loop, training local community members on the monitoring tools needed and thus empowering them (Björkman Nyqvist, de Walque, and Svensson 2017). Furthermore, as reflected by the political turmoil around the SAM in Kondoa and other evidence (Mamdani et al., 2018), Sikika's reputation of publicly exposing wrongdoing among public officials likely triggered improved responsiveness among district authorities. Our interpretation is also consistent with stream of literature suggesting that the ability to maintain independence from formal government mechanisms is critical for SAM programs (Feruglio & Nisbett, 2018). This seems to be particularly relevant when monitoring programs are embedded in formal government structures or when citizens are recruited for monitoring policies without proper training. These findings are supported by qualitative data purposely collected in the two treatment districts of Kondoa and Mpwapwa (see Appendix 14 for further details). Overall, focus group discussions and interviews suggest that both health facility workers and citizens appreciated the SAM program. The regular meetings appear to have resulted in increased feedback to facilities, and also may have increased both the direct social pressure from the community to perform well and indirect pressure through exchanges with the district office.

Even though the SAM program studied achieved only some of its objectives, we believe the program did have a meaningful impact on a crucial element for the functioning of health systems. Access to essential medicines is among the top global health priorities, as reflected by the United Nations Sustainable Development Goals (goal 8E, specifically). Despite major efforts in the past decades, making essential medicines available in a consistent and reliable fashion was and remains a major challenge in many countries due to several and frequently concurrent inefficiencies in healthcare delivery. Stockouts of essential medicines have been linked to gaps in immunization coverage (Favin, Robert SteinglassFields, Banerjee, & Sawhney, 2012), poor control of noncommunicable diseases (Attaei et al., 2017), ineffective antiretroviral therapy (Berheto, Haile, and Mohammed 2014), incomplete detection and treatment of malaria (Layer et al., 2014), and increased maternal and child mortality (Githinji et al., 2013). Stockouts of essential medicines at public health facilities also have been shown to increase out-of-pocket payments by forcing patients to

purchase required drugs from private providers (Mikkelsen-Lopez, Shango, Jim BarringtonZiegler, Smith, & Don, 2014; Wagenaar et al., 2014; Wales, Tobias, Malangalila, Godfrey, & Wild, 2014), to generate dissatisfaction with services and to inhibit health seeking behavior (Ikoh, Udo, Charles, & Charles, 2009; Kruk, Rockers, Godfrey, Paczkowski, & Galea, 2010; Muhamadi et al., 2010; Tefera, Tesfaye, Abeba BekeleElias, Waltensperger, & Marsh, 2014). Lastly, low availability of essential medicines has been associated with low enrollment in social or community-based health insurance schemes (Fadlallah et al., 2018; Kalolo et al., 2015; Kamuzora & Gilson, 2007; Renggli et al., 2019).

Our study has important implications for LMICs planning to introduce social accountability mechanisms as tools to improve bottom-up governance in the health system. The lack of effect on outcomes related to infrastructure and the differential results across drug classes suggests that the objectives and targets of SAM programs should be carefully reviewed and locally adapted. These results indicate that SAM programs are effective in boosting the performance of health facilities where functioning local health systems are in place. This implies that social accountability alone may not produce the expected results in settings where broader structural health systems failures hinder the performance of health facilities. Furthermore, social accountability seems more likely to yield the desired results in settings where changes in provider efforts yield tangible results. Areas of health care characterized by binding institutional constraints (e.g., drug delivered through vertical programs, investments in infrastructure requiring long application for additional funds, etc.) are unlikely to respond to increased provider efforts. Targeting outcomes in areas that cannot easily be controlled by health workers thus seems somewhat unlikely to yield positive results and should be avoided in the design of future SAM programs. Finally, the results of this study are consistent with those from previous literature, suggesting that social accountability initiatives should go beyond simple information sharing and include a strong monitoring component. This entails promoting feedback loops between community, service providers and authorities, opening channels for the government to better react to citizens' "voice" (Fox, 2015; Ringold et al., 2012), as well as building capacity, providing tools and training citizens' involved, enabling them to fulfill their roles (Lopez Franco & Shankland 2018).

CONFLICT OF INTEREST STATEMENT
The authors report no conflicts of interest.

ETHICAL APPROVAL
Ethical clearance for this study was obtained from the Ifakara Health Institute (IHI) Institutional Review Board (IHI/IRB/No: 20-2017) and the Tanzanian National Institute for Medical Research (NIMR/HQ/R.8a/Vol.IX/2721).

AUTHORS CONTRIBUTION
Igor Francetic developed the analytical framework, prepared the data, performed the data analysis and took the lead in writing the manuscript. Günther Fink and Fabrizio Tediosi provided critical feedback and helped shape the research, analysis and manuscript. All authors approved the final version of the manuscript.

DATA AVAILABILITY STATEMENT
Data available on request from the authors.

ORCID
Igor Francetic https://orcid.org/0000-0002-2481-3749

ENDNOTES

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

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

REFERENCES

Adhvaryu, A., & Nyshadham, A. (2015). Return to treatment in the formal health care sector: Evidence from Tanzania. *American Economic Journal: Economic Policy*, *7*(3), 29–57. https://doi.org/10.1257/pol.20120262

Angrist, J. D., & Pischke, J. -S. (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton, USA: Princeton University Press.

Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *The Journal of Economic Perspectives*, *31*(2), 3–32. https://doi.org/10.1257/jep.31.2.3

Attaei, M. W., Khatib, R., McKee, M., Lear, S., Dagenais, G., Igumbor, E., ... & 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*(9), e411–e419. https://doi.org/10.1016/S2468-2667(17)30141-X

Berheto, Moshago, T., Demissew Berihun, H., & 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*(9), 453–459. https://doi.org/10.4103/1947-2714.141636

Binyaruka, P., & Borghi, J. (2017). Improving quality of care through payment for performance: Examining effects on the availability and stock-out of essential medical commodities in Tanzania. *Tropical Medicine and International Health*, *22*(1), 92–102. https://doi.org/10.1111/tmi.12809

Björkman, M., & Svensson, J. (2009). Power to the people: Evidence from a randomized field experiment on community-based monitoring in Uganda. *Quarterly Journal of Economics*, *124*(2), 735–769. https://doi.org/10.1162/qjec.2009.124.2.735

Björkman, N., Martina, D. de W., & Svensson, J. (2017). Experimental evidence on the long-run impact of community-based monitoring. *American Economic Journal: Applied Economics*, *9*(1), 33–69. https://doi.org/10.1257/app.20150027

Boex, J., Fuller, L., & Malik, A. (2015). *Decentralized local health services in Tanzania.* Washington, DC: Urban Institute.

Boydell, V., McMullen, H., Cordero, J., Steyn, P., & James, K. (2019). Studying social accountability in the context of health system strengthening: Innovations and considerations for future work. *Health Research Policy and Systems*, *17*(1), 34. https://doi.org/10.1186/s12961-019-0438-x

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, *90*(3), 414–427. https://doi.org/10.1162/rest.90.3.414

Danhoundo, G., Nasiri, K., Mary, E., & Wiktorowicz (2018). Improving social accountability processes in the health sector in sub-Saharan Africa: A systematic Review. *BMC Public Health*, *18*(1), 497. https://doi.org/10.1186/s12889-018-5407-8

Dimick, J. B., & Ryan, A. M. (2014). Methods for evaluating changes in health care policy: The difference-in-differences approach. *Journal of the American Medical Association*, *312*(22), 2401–2402. https://doi.org/10.1001/jama.2014.16153

Ding, P., & Fan, Li (2019). A bracketing relationship between difference-in-differences and lagged-dependent-variable adjustment. *Political Analysis*, *27*(4), 605–615. https://doi.org/10.1017/pan.2019.25

Donald, Stephen, G., & Lang, K. (2007). Inference with difference-in-differences and other panel data. *The Review of Economics and Statistics*, *89*(2), 221–233. https://doi.org/10.1162/rest.89.2.221

Egedesø, P. J., Casper Worm, H., & Peter Sandholt, J. (2020). Preventing the white death: Tuberculosis dispensaries. *The Economic Journal*, *130*(629), 1288–1316. https://doi.org/10.1093/ej/ueaa014

Fadlallah, R., El-Jardali, F., Hemadi, N., Morsi, R. Z., Clara Abou, A. S., Ahmad, A., ... Akl (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*(1), 13. https://doi.org/10.1186/s12939-018-0721-4

Favin, M., Robert Steinglass, Fields, R., Banerjee, K., & Sawhney, M. (2012). Why children are not vaccinated: A Review of the grey literature. *International Health*, *4*(4), 229–238. https://doi.org/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*(1), 788. https://doi.org/10.1186/s12913-018-3600-1

Fiala, N. V., & Premand, P. (2018). *Social accountability and service delivery: Experimental evidence from Uganda.* Washington, DC: World Bank. https://doi.org/10.1596/1813-9450-8449

Fox, J. (2007). The uncertain relationship between transparency and accountability. *Development in Practice*, *17*(4–5), 663–671. https://doi.org/10.1080/09614520701469955

Fox, J. (2015). Social accountability: What does the evidence really say? *World Development*, *72*, 346–361. https://doi.org/10.1016/j.worlddev.2015.03.011

Gaventa, J., & McGee, R. (2013). The impact of transparency and accountability initiatives. *Development Policy Review*, *31*(s1), s3–s28. https://doi.org/10.1111/dpr.12017

Githinji, S., Kigen, S., Memusi, D., Nyandigisi, A., Mbithi, A. M., Wamari, A., ... 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*(1), e54066. https://doi.org/10.1371/journal.pone.0054066

Greer, S. L., Matthias, W., & Josep, F., & European Observatory on Health Systems and Policies. (2016). *Strengthening health system governance: Better policies, stronger performance. European observatory on health systems and policies series.* Maidenhead, Berkshire, UK: Open University Press.

Hamal, M., Tjard de Cock, B., 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*(1), 653. https://doi.org/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'. *The International Quarterly of Community Health Education*, *29*(3), 257–273. https://doi.org/10.2190/IQ.29.3.e

Joshi, A. (2017). Legal empowerment and social accountability: Complementary strategies toward rights-based development in health? *World Development*, *99*(November), 160–172. https://doi.org/10.1016/j.worlddev.2017.07.008

Joshi, A., & Houtzager, P. P. (2012). Widgets or watchdogs? *Public Management Review*, *14*(2), 145–162. https://doi.org/10.1080/14719037.2012.657837

Kalolo, A., Radermacher, R., Stoermer, M., Meshack, M., & De Allegri, M. (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*(1), 29648. https://doi.org/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*(2), 95–102. https://doi.org/10.1093/heapol/czm001

Kantor, D. (2012). *MAHAPICK: Stata Module to select matching observations Based on a Mahalanobis distance measure. Statistical software components*. Retrieved from https://econpapers.repec.org/software/bocbocode/s456703.htm

Kigume, R., & Maluka, S. (2018). Decentralisation and health services delivery in 4 districts in Tanzania: How and why does the use of decision space vary across districts? *International Journal of Health Policy and Management*, *8*(2), 90–100. https://doi.org/10.15171/ijhpm.2018.97

Kruk, M. E., Rockers, P. C., Godfrey, M., Paczkowski, M. M., & Galea, S. (2010). Community and health system factors associated with facility delivery in rural Tanzania: A multilevel analysis. *Health Policy*, *97*(2), 209–216. https://doi.org/10.1016/j.healthpol.2010.05.002

Kuwawenaruwa, A., Wyss, K., Wiedenmayer, K., Metta, E., & Tediosi, F. (2020). 'The effects of medicines availability and stock-outs on household's utilization of healthcare services in Dodoma region, Tanzania'. *Health Policy and Planning*, *35*(3), 323–333. https://doi.org/10.1093/heapol/czz173

Layer, E. H., Kennedy, C. E., Beckham, S. W., Mbwambo, J. K., Likindikoki, S., Wendy, W., ... Kerrigan, H. B., & the LTC Tanzania Collaborative Study Team. (2014). Multi-level factors affecting entry into and engagement in the HIV continuum of care in iringa, Tanzania. *PloS One*, *9*(8), e104961. https://doi.org/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*(1), 19–35. https://doi.org/10.1080/13645579.2018.1488449

Lewis, M. (2006). *Governance and corruption in public health care systems* SSRN scholarly paper ID 984046. Center for global development working paper No. 78. Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=984046

Lewis, M., & Pettersson, G. (2009). *Governance in health care Delivery : Raising performance* Policy research working papers. The World Bank. Retrieved from https://doi.org/10.1596/1813-9450-5074

Lodenstein, E., Dieleman, M., Gerretsen, B., & Broerse, J. E. W. (2017). Health provider responsiveness to social accountability initiatives in low- and middle-income countries: A realist Review. *Health Policy and Planning*, *32*(1), 125–140. https://doi.org/10.1093/heapol/czw089

Lodenstein, E., Ingemann, C., Molenaar, J. M., Dieleman, M., Jacqueline, E., & Broerse, W. (2018). Informal social accountability in maternal health service delivery: A study in northern Malawi. *PloS One*, *13*(4), e0195671. https://doi.org/10.1371/journal.pone.0195671

Lopez Franco, E., & Shankland, A. (2018). *Guidelines for Designing and Monitoring Social Accountability Interventions*, Brighton: SDC-IDS. https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/14105

MacKinnon, J. G., & Webb, M. D. (2018). The wild bootstrap for few (treated) clusters. *The Econometrics Journal*, *21*(2), 114–135. https://doi.org/10.1111/ectj.12107

Mamdani, M., Kweka, H., Peter, B., Ramesh, M., Kapologwe, N., Hutchinson, E., ... Andreoni, A. (2018). *Strengthening accountability for better health outcomes through understanding healthsystem bottlenecks: Insights from Tanzania* Working paper 008. ACE working paper. London, UK: Anti-Corruption Evidence SOAS Consortium.

Martin, H., Adriane, C. B., 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*(3), 345–357. https://doi.org/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*(6), 449–466. https://doi.org/10.1093/heapol/czr077

Mikkelsen-Lopez, I., Shango, W., Jim Barrington, Ziegler, R., Smith, T., & Don, deS. (2014). The challenge to avoid anti-malarial medicine stock-outs in an era of funding partners: The case of Tanzania. *Malaria Journal*, *13*(1), 181. https://doi.org/10.1186/1475-2875-13-181

Ministry of Health and Social Welfare. (2011). *Annual health statistical tables and figures Tanzania Mainland. HMIS 2011. Dodoma: The United Republic of Tanzania*.

Ministry of Health, & Community Development, Gender, Elderly and Children. (2019). *Health facility registry (HFR). Dodoma: The United Republic of Tanzania*.

Moïsi, J. C., Gatakaa, H., Noor, A. M., Williams, T. N., Bauni, E., Benjamin, T., ... 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*(1), 142. https://doi.org/10.1186/1471-2458-10-142

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*(4), 462–499. https://doi.org/10.1080/19439342.2017.1378243

Muhamadi, L., Nsabagasani, X., Mbona Nazarius, T., Wabwire-Mangen, F., Ekström, A.-M., Peterson, S., & George, P. (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*(2), 187–194. https://doi.org/10.1016/j.healthpol.2010.06.003

Musau, S., Chee, G., Patsika, R., Malangalila, E., Chitama, D., Van Praag, E., & Schettler, G. (2011). *Tanzania health system Assessment 2010 report*. Bethesda, MD: Health Systems 20/20 project, Abt Associates Inc.

National Bureau of Statistics, & Office of Chief Government Statistician. (2013). *2012 popopation and housing census. Population distribution by administrative areas*. Dar es Salaam, Tanzania: United Republic of 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*(1), 1421002. https://doi.org/10.1080/16549716.2017.1421002

O'Meally, S. C. (2013). *'Mapping context for social accountability: A resource paper'. 79351*. Washington DC, USA: World Bank.

O'Neill, S., Kreif, N., Grieve, R., Sutton, M., Jasjeet, S., & Sekhon. (2016). Estimating causal effects: Considering three alternatives to difference-in-differences estimation. *Health Services & Outcomes Research Methodology*, *16*, 1–21. https://doi.org/10.1007/s10742-016-0146-8

Pelt, S. van, Massar, K., van der Eem, L., Laura Shields-Zeeman, J. B., de Wit, F., Robert, A., & Ruiter, C. (2020). "If you don't have enough equipment, you're not going to provide quality services": Healthcare workers' perceptions on improving the quality of antenatal care in rural Tanzania. *International Journal of Africa Nursing Sciences*, *13*(January), 100232. https://doi.org/10.1016/j.ijans.2020.100232

Penfold, S., Shamba, D., Hanson, C., Jaribu, J., Manzi, F., Marchant, T., ... Armstrong Schellenberg, J. (2013). 'Staff experiences of providing maternity services in rural southern Tanzania – a focus on equipment, drug and supply issues'. *BMC Health Services Research*, *13*(1), 61. https://doi.org/10.1186/1472-6963-13-61

Raza, W. A., Ellen Van de Poel, & Van Ourti, T. (2018). Impact and spill-over effects of an asset transfer program on child undernutrition: Evidence from a randomized control trial in Bangladesh. *Journal of Health Economics*, *62*, 105–120.

Renggli, S., Mayumana, I., Mshana, C., Mboya, D., Kessy, F., Tediosi, F., ... 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*, *34*(1), 12–23. https://doi.org/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*. Washington, DC: World Bank.

Rokicki, S., Cohen, J., Fink, G. ., Salomon, J. A., & Beth Landrum, M. (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*(1), 97–105. https://doi.org/10.1097/MLR.0000000000000830

Scholz, S., Ngoli, B., & Flessa, S. (2015). Rapid assessment of infrastructure of primary health care facilities – a relevant instrument for health care systems management. *BMC Health Services Research*, *15*(1), 183. https://doi.org/10.1186/s12913-015-0838-8

Sikika. (2013). *Annual Report 2012*. Retrieved from http://sikika.or.tz/reports/ Dodoma: Sikika. https://sikika.or.tz/index.php/en/narrative-reports/annual-report-2012/download

Sikika. (2014). *Annual Report 2013*. Dodoma: Sikika. https://sikika.or.tz/index.php/en/narrative-reports/annual-report-2014/download

Sikika. (2015). *Annual Report 2014*. Dodoma: Sikika. https://sikika.or.tz/index.php/en/narrative-reports/annual-report-2014/download

Sikika. (2016a). Ripoti ya ufuatiliaji Na uwajibikaji Kwa jamii (SAM). In *Halmashauri Ya Wilaya Ya Kondoa Vijijini. Imeandaliwa Na Timu Ya SAM* Septemba 2016. Dodoma, Tanzania: Sikika.

Sikika. (2016b). *Annual Report 2015*. Dodoma: Sikika. https://sikika.or.tz/index.php/en/narrative-reports/annual-report-2015/download

Stoermer, M., Werlein, R., & Molesworth, K. (2011). *Health promotion and system strengthening (HPSS) project in Dodoma region. Tanzania*. Project document. Basel, Switzerland: Swiss Centre for International Health.

Sumra, S., Ruto, S., Rajani, R., & Gaddis, I. (2015). Assessing literacy and numeracy in Tanzania's primary schools: The uwezo approach. In A. R. Joshi (Ed.), *Preparing the next generation in Tanzania: Challenges and opportunities in education* (pp. 47–64) The World Bank. https://doi.org/10.1596/978-1-4648-0590-5_ch4

Tefera, W., Tesfaye, H., Abeba Bekele, Elias, K., 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.

University Consultancy Bureau. (2018). *Health facility survey report for A follow up survey for the health promotion and system strengthening (HPSS)*. Dar es Salaam, Tanzania: University of Dar es Salaam.

Wagenaar, B. H., Gimbel, S., Hoek, R., Pfeiffer, J., Michel, C., Luis Manuel, J., ... Sherr, K. (2014). 'Stock-outs of essential health products in Mozambique – longitudinal analyses from 2011 to 2013'. *Tropical Medicine and International Health*, *19*(7), 791–801. https://doi.org/10.1111/tmi.12314

Wales, J., Tobias, J., Malangalila, E., Godfrey, S., & Wild, L. (2014). Stock-Outs of essential medicines in Tanzania. In *A political economy approach to analysing problems and identifying solutions*. London, UK: Overseas Development Institute.

Webb, M. D. (2013). *Reworking wild bootstrap based inference for clustered errors* Working Paper 1315.

Wiedenmayer, K. (2017). *JAZIA Prime vendor system Dodoma region*. Basel, Switzerland: Swiss Tropical and Public Health Institute (Swiss TPH).

Wiedenmayer, K., Mbwasi, R., Mfuko, W., Mpuya, E., Charles, J., & Chilunda, F. (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*(1), 4. https://doi.org/10.1186/s40545-019-0163-4

Wooldridge, J. M. (Ed). (2010). *Econometric analysis of cross section and panel data* (2nd ed.). Cambridge, MA: MIT Press.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Francetic I, Fink G, Tediosi F. Impact of social accountability monitoring on health facility performance: Evidence from Tanzania. *Health Economics*. 2021;1−20. https://doi.org/10.1002/hec.4219