NBER WORKING PAPER SERIES

THE GOVERNANCE OF NON-PROFITS AND THEIR SOCIAL IMPACT: EVIDENCE FROM A RANDOMIZED PROGRAM IN HEALTHCARE IN THE DEMOCRATIC REPUBLIC OF CONGO

Anicet Fangwa
Caroline Flammer
Marieke Huysentruyt
Bertrand Quelin

Working Paper 30391
http://www.nber.org/papers/w30391

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2022

We are grateful to Ruth Aguilera, Dominique Baabo Kubuya, Marco Becht, Vanessa Burbano, Olga Hawn, Natasha Hochberg, Steve Lydenberg, Serge Mayaka, Mary-Hunter McDonnell, Agnes Muzama, Lamar Pierce, Ivan Png, Gui Deng Say, David Stolow, Haitao Yu, as well as seminar participants at Boston University, Georgia Institute of Technology, George Mason University, HEC Paris, IESE, Ivey, Lancaster University, Singapore Management University, Toulouse School of Economics, UCLA, University of Manchester, University of Zurich, UT Austin, the Ackerman Corporate Governance Conference (Tel Aviv), the Swiss Economists Abroad Conference (Geneva), and the Strategic Management Society Annual Conference (Minnesota) for helpful comments and suggestions. We thank the DRC's Ministry of Health, the PDSS coordination team, and Bluesquare for providing us with the data as well as their help and insights. This research is supported by a grant of the French National Research Agency (ANR) "Investissements d'Avenir" (LabEx Ecodec). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Governance of Non-Profits and their Social Impact: Evidence from a Randomized Program in Healthcare in the Democratic Republic of Congo

Anicet Fangwa, Caroline Flammer, Marieke Huysentruyt, and Bertrand Quelin

NBER Working Paper No. 30391
August 2022
JEL No. I0,I1,I2,O1

ABSTRACT

How can non-profit organizations improve their governance to increase their social impact? This study examines the effectiveness of a bundle of governance mechanisms – consisting of social performance-based incentives combined with auditing and feedback – in the context of a randomized governance program conducted in the Democratic Republic of Congo's healthcare sector. Within the program, a set of health centers were randomly assigned to a governance treatment while others were not. We find that the governance treatment leads to i) higher operating efficiency and ii) improvements in health outcomes (measured by a reduction in stillbirths and neonatal deaths). Furthermore, we find that funding is not a substitute for governance – health centers that only receive funding increase their scale, but do not show improvements in operating efficiency nor health outcomes. Overall, our results suggest that governance plays an important role in achieving the non-profits' objectives and increasing the social impact of the funds invested.

Anicet Fangwa
HEC Paris
1 Rue de la Libération
78350 Jouy-en-Josas
France
anicet.fangwa@hec.edu

Marieke Huysentruyt
HEC Paris
1 Rue de la Libération
78350 Jouy-en-Josas
France
huysentruyt@hec.fr

Caroline Flammer
School of International and Public Affairs
Columbia University
420 West 118th Street, Office 1429
New York, NY 10026
and NBER
caroline.flammer@columbia.edu

Bertrand Quelin
HEC Paris
1 Rue de la Libération
78350 Jouy-en-Josas
France
quelin@hec.fr

A data appendix is available at http://www.nber.org/data-appendix/w30391

1. Introduction

Every year, considerable efforts and large amounts of funds are invested in social and environmental causes. Those include ending poverty, reducing inequalities, promoting healthy lives, addressing climate change, and protecting the world's biodiversity, among others. Accounting for about one-third of total employment in the social sector, non-profit organizations represent a large part of the global economy and constitute a major player in pursuing these efforts (United Nations 2018a).

To help tackle these grand societal challenges, it is important to understand how to improve the effectiveness of non-profit organizations. Arguably, this is a critical question not only for academics but also, and more importantly, for the non-profit organizations themselves and the beneficiaries of their services—affected individuals, communities, and the natural environment. What is more, ensuring the efficient use of funds is also important for donors and impact investors who aim to make a difference in this world and maximize their funds' social impact. In sum, understanding what mechanisms are available and effective in improving the social impact of non-profit organizations is important for the organizations per se, the donors and impact investors, as well as society at large and the natural environment.

The question of how to improve the governance of non-profits is difficult to answer, both theoretically and empirically. The insights from the existing literature on governance offer only limited guidance in this regard since it has largely focused its attention on for-profit organizations, which are fundamentally different from non-profit organizations.[1]

From a theoretical perspective, it is unclear what governance mechanisms are available and effective for non-profits to maximize the social impact of the funds invested, as well as when and where they matter more or less. In order to be effective, appropriate governance mechanisms need to take into

[1] By their very nature, non-profits pursue social as opposed to financial objectives. In this regard, an inherent feature of non-profits is the "non-distribution constraint" (Hansmann 1980)—that is, non-profits are not allowed to distribute profits to donors or employees; instead, any surplus they generate must be retained and devoted to their social objectives. A direct implication of this non-distribution constraint is that non-profits do not have owners. The investors who fund non-profits, through donations, do not have any claim to the non-profits' revenues and assets, nor do they have any control rights over the organization. Accordingly, many of the governance tools available to for-profits (e.g., managerial ownership) are not available to non-profits.

account the unique nature of and obstacles faced by non-profit organizations. First, non-profits are often cash-constrained, their employees might be purpose-driven and intrinsically motivated, and key (non-financial) performance metrics might be difficult to measure and evaluate. Second, many non-profits operate in countries that are resource-constrained (e.g., in terms of financial resources and qualified personnel). In this regard, the provision of additional funding is likely helpful. However, funding per se need not be sufficient to bring about tangible improvements in the non-profits' social impact. In particular, their managers and employees may lack the necessary knowledge on how to improve the organization's operating efficiency and service quality. Accordingly, even if they had access to more funding, they might not be able to put this funding to good use without better governance practices. For example, in the context of the health sector in the Democratic Republic of Congo (DRC), the WHO has identified several sources of inefficiencies such as the lack of strategic and managerial planning, inadequate priorities in resource allocation, lack of transparency, lack of managerial competencies, and insufficient medical training of health workers (WHO 2015). These sources of inefficiencies are unlikely unique to the health sector in the DRC, but rather a common challenge found across sectors and across (low-income) countries. Given these challenges, it is unclear whether and to what extent the provision of funding and better governance can truly move the needle and bring about substantial improvements in operating efficiency and service quality.

From an empirical perspective, making ground on these questions is challenging. There are two main obstacles. First, it is difficult to obtain fine-grained microdata on non-profit organizations, their governance, as well as their social impact. Second, even if detailed microdata were available, the adoption of governance practices is likely endogenous with respect to organizational outcomes—that is, unobservables may drive a spurious relationship between the adoption of different governance mechanisms and organizational outcomes. Addressing the endogeneity of governance requires a source of exogenous variation in the adoption of governance practices.

To overcome these obstacles, we study the governance of non-profits in the context of a randomized governance program implemented in the healthcare sector of the Democratic Republic of Congo (DRC) by the World Bank. In the DRC, primary healthcare services—and especially maternity and childbirth

services—are administered in non-profit health centers spread across the country. The randomized governance program we exploit in this paper was administered in about 1,000 health centers at the beginning of 2017. Health centers in the program were randomly assigned to a treatment group and control group, respectively. While health centers in both groups received funding from the program, only those in the treatment group were subject to a "governance treatment" consisting of social performance-based incentives combined with auditing and feedback (A&F).[2] Hence, by design, this randomized governance program provides an ideal setup to study how the adoption of specific governance practices affects health centers' outcomes (e.g., their operating efficiency and service quality), holding everything else, including funding, constant.

Note that we refer to the intervention as a "governance" treatment, since the intervention explicitly aims to align the interests of the health center's staff toward the health center's objective, that is, the common social goal of improving the quality and quantity of health services.[3] That said, while the components of the intervention can be seen as governance practices, they can also be construed as management practices, as they directly affect the way the health center's daily operations are conducted.[4]

Using a difference-in-differences methodology, we find that, within a ten-quarter period following the treatment, health centers in the treatment group experience significant improvements in both operating efficiency (captured by an increase in the number of services provided per employee) and health outcomes (captured by a reduction in the probability of stillbirths and neonatal deaths, respectively). These findings suggest that the adoption of specific governance mechanisms (in the form of social performance-based incentives and A&F) is beneficial to non-profits and contributes to their ability to achieve their intended

[2] Social performance-based incentives are a form of pay-for-social-performance. Specifically, additional funding is provided to the non-profit organization contingent on the organization's social performance. See Section 3.2 for details.

[3] This is in line with the way governance is typically defined in the literature. For example, in their textbook, Hanson et al. (2017, p. 292) define governance as "the set of mechanisms used to manage the relationship among stakeholders and to determine and control the strategic direction and performance of organisations."

[4] Similarly, in their discussion of the role of governance vs. management practices in the healthcare sector, the Governance Institute (2018) notes that "[t]he line between governance and management can sometimes blur. In fact, we believe there is no real 'line' between the two" (p. 1).

social impact.

To shed light on the mechanism, we conduct a series of interviews with program participants. Collectively, the respondents highlighted that a key challenge was the lack of adequate training and organizational know-how. In this regard, the quarterly rounds of feedback—combined with incentives to actually act on this feedback—were seen as essential in inducing tangible changes that would ultimately translate into higher efficiency and quality of services. Overall, our interviews suggest that combining learning with incentives to implement what is learnt is a plausible mechanism underlying our findings.[5]

In addition, to gain insights into what happened within the "black box" of the treated health centers, we use finer-grained data that track the evolution of several quality scores that were collected as part of the PDSS program. Three patterns emerge from this analysis. First, the largest improvements are found in the categories planning, financial management, and general organization. These improvements provide a plausible rationale for the increase in operating efficiency. Second, we also observe large improvements in the categories maternity care, pre-natal care, and family planning. These categories are all directly related to maternal and childhood health. As such, they potentially explain the reduction in infant mortality. Third, these improvements tend to materialize within the initial quarters of the PDSS program and remain somewhat stable thereafter. This is again consistent with the learning interpretation, in which the initial quarters of the program are essentially a learning phase that subsequently translates into higher operating efficiency and lower infant mortality.

Finally, we examine whether funding can serve as a substitute for governance. This question is not only important from a societal perspective but also from an implementation and policy perspective. Indeed, transferring money is relatively straightforward, while it is more challenging to implement better governance and management practices. To shed light on this question, we compare health centers in- and

[5] In auxiliary analyses, we find further evidence that points toward the importance of learning. Specifically, we find that the treatment effect is stronger in areas with a lower density of health centers (that is, areas where it is harder to learn and adopt best practices from their peers). Moreover, we find that i) the increase in operating efficiency is higher for health centers with lower ex ante operating efficiency, and ii) the decrease in infant mortality is more pronounced for health centers with higher ex ante mortality. Taken together, these findings indicate that the improvements brought about by the treatment are larger when there is more to improve and learn to begin with.

outside the governance program. We refer to the latter group as the "outside group." Unlike the treatment group (i.e., health centers that receive both funding and the governance treatment) and the control group (i.e., health centers that only receive funding), health centers in the outside group receive neither. Accordingly, by comparing the treatment group versus the outside group, we can estimate the impact of the combination of 'funding and governance' on health center outcomes. Similarly, by comparing the control group versus the outside group, we can measure the impact of 'funding' as standalone. This analysis reveals that funding is not a substitute for governance—health centers that only receive funding increase their scale, but do not show improvements in operating efficiency nor service quality. In contrast, health centers that receive both funding and the governance treatment improve both their scale as well as their operating efficiency and service quality.

Overall, our results suggest that governance plays an important role in contributing to non-profits' objectives and increasing their social impact, and that funding alone does not serve as a substitute for governance. Naturally, we caution that our findings are specific to the healthcare sector in a low-income country, and hence need not generalize to the overall non-profit sector. Nevertheless, studying this specific context is important, as improving children's health has been an important challenge in economically developing countries. In particular, according to the World Health Organization (WHO), the mortality rate of children under the age of five decreased from 9.3% (i.e., 93 deaths per 1,000 live births) in 1990, to 3.9% in 2017 (WHO 2018). This corresponds to a worldwide decline from 12.6 million deaths in 1990, to 5.4 million in 2017. Yet, despite this overall progress, stark disparities exist across regions and countries. In particular, Sub-Saharan Africa remains the region with the highest under-5 mortality rate in the world, with one child in thirteen dying before their fifth birthday—this is fourteen times higher than in high-income countries (WHO 2018). A similar picture arises with the mortality rate of infants under one-year old. As Figure A1 shows, the DRC is among the countries with the highest infant mortality rates in the world. An estimated 7% of infants died within the first year of their life in 2017, compared to 0.4% in France and 0.6% in the U.S. (United Nations 2018b). Reducing children mortality is of foremost importance, and this study helps inform how specific governance instruments can contribute to this objective.

2. The role of governance in improving non-profits' operating efficiency and their social impact

In the following, we discuss the potential effectiveness of a bundle of governance mechanisms—consisting of social performance-based incentives combined with auditing and feedback (A&F)—for improving the non-profit organizations' operating efficiency and social performance. This governance bundle is used in the randomized governance program we study in the empirical analysis.

Social performance-based incentives

To motivate managers and employees to use the non-profits' resources efficiently and maximize their social impact, donors (and other impact investors) can provide social performance-based incentives. That is, additional funding is provided to the non-profit organization conditional on meeting specific social criteria. As such, social performance-based incentives differ from the more traditional financial incentives in two ways: i) the additional funding is tied to social criteria in lieu of financial criteria, and ii) the direct beneficiary of social performance-based incentives is the non-profit organization itself as opposed to the managers and employees, respectively.

Social performance-based incentives may motivate non-profit managers and employees in two ways. First, by providing such incentives, non-profits can leverage managers' and employees' intrinsic motivation to obtain additional funding for the non-profit's cause.[6] Second, social performance-based incentives may provide benefits to the managers and employees. Indeed, the additional funding allows non-profit organizations to, e.g., upgrade their equipment, provide training to their employees, pay higher wages, extend the scope of their services, all of which increase the attractiveness of the workplace and, as a result, can have a motivational effect on individuals' work behavior.[7] On top of the motivational aspect, providing

[6] Relatedly, recent studies in the context of for-profits in high-income countries have shown that employees across various occupations (e.g., consultants, lawyers, and online workers) are willing to forgo financial compensation for the pursuit of "meaningful" work (e.g., Bode and Singh 2018, Burbano 2016, Carnahan, Kryscynski, and Olson 2017, Cassar 2019, Cassar and Meier 2018).

[7] In a similar spirit, non-financial benefits awarded to employees (e.g., through employer recognition, social visibility, or employee satisfaction programs) are found to motivate employees and improve their productivity (e.g., Ashraf, Bandiera, and Jack 2014, Ashraf, Bandiera, and Lee 2014, Dur, Non, and Roelfsema 2010, Flammer and Luo 2017, Gallus and Frey 2016, Gubler, Larkin, and Pierce 2017).

employees and managers with training, upgraded equipment, etc., may empower them to further increase their productivity.

While the provision of social performance-based incentives may positively affect the motivation and productivity of the non-profits' managers and employees, there are also several reasons as to why they may not be effective, or even backfire.

First, employees of non-profits are likely to be purpose-driven and intrinsically motivated. In this regard, the provision of performance-based incentives could crowd out their intrinsic motivation and ultimately reduce their engagement and job satisfaction.[8] In addition to decreasing employees' intrinsic motivation, such incentives may reduce performance if the employees lack the necessary knowledge on how to improve (Huillery and Seban 2021). Second, compared to financial performance metrics, non-financial performance metrics are difficult to measure and hence likely less suitable as a basis for incentive schemes.[9] Moreover, the difficulty of measuring, tracking, and verifying non-financial performance might increase the risk of manipulating performance records to increase payouts (Linden and Shastry 2012). Third, in the presence of multi-tasking activities, performance incentives may encourage substitution away from non-incentivized activities, reorienting the employees' attention toward those activities with a greater pay-for-performance component or/and those that require lower effort (Basinga et al. 2011, Holmstrom and Milgrom 1991).

Auditing and feedback (A&F)

In addition to social performance-based incentives, A&F can be introduced to enhance the non-profits' operating efficiency and social performance. Specifically, by conducting audits on a regular basis, independent third parties can verify that best practices and protocols are followed, adequate priorities are set in terms of resource allocation, a strategic and managerial plan is developed and pursued, and that the

[8] This argument echoes the findings of studies in the context of for-profit organizations and the provision of financial performance-based incentives (e.g., Bowles 2016, Cassar and Meier 2021, Gubler, Larkin, and Pierce 2016, Wrzesniewski et al. 2014).

[9] In the for-profit context, pay-for-social performance incentive schemes have been shown to be ineffective when they are vague and insufficiently specified (Flammer, Hong, and Minor 2019).

organization's practices and performance are correctly documented.[10]

Through such audits, the non-profits also obtain data-based feedback on discrepancies between current and target performance. The recognition of underperformance may encourage them to take action and improve their practices. Moreover, given their business expertise, auditors can play an important role in providing feedback and recommendations to the non-profits' managers and employees, thereby acting as informal coaches.[11] Such coaching might be of foremost importance. Indeed, an inherent challenge faced by non-profits in low-income countries is to find competent managers and well-trained personnel. The lack of know-how can lead to, e.g., unrealistic planning, inefficient work deployment, inadequate priorities in resource allocation, improper interpretation and implementation of guidelines and procedures, and overall poor decision-making; all of which are likely to result in the inefficient use of the non-profit's already limited financial resources, inventory, and human capital. Accordingly, by sharing valuable insights, auditors can help transfer best practices and improve the non-profits' effectiveness in terms of both their operating efficiency and quality of services. Moreover, feedback can be reassuring for individuals and help them build up their self-confidence that they are doing things right (Bandiera et al. 2015).

On the other hand, A&F may also have detrimental effects on employees. In particular, rating employees may reduce their self-esteem and create animosity and competition among peers (Smither et al. 2005). Relatedly, managers' evaluation of subordinates (which is commonly done in traditional performance evaluation and in "rank-and-yank" performance appraisals) has been found to crowd out creativity, as employees may hesitate to suggest unusual, creative ideas for fear of a harsh evaluation (Amabile et al. 1990).

Financial subsidies

In addition to governance practices, financial subsidies can play an important role in enhancing the social

[10] For example, in the context of the DRC's health centers examined in this study, regular visits of independent auditors can ensure that general guidelines provided by the WHO are followed (PDSS 2016).
[11] For instance, auditors of the DRC's health centers examined in this study provide feedback to the audited health centers in order to improve their operations and help them follow best practices (PDSS 2016).

impact of non-profits, especially in resource-constrained environments. The provision of financial subsidies helps alleviate financing constraints and allows non-profits to invest in, e.g., upgrading their equipment, hiring more employees, and extending the scope of their services. This likely increases the scale of the non-profits' operations. Thus, the provision of financial subsidies in combination with the implementation of governance mechanisms may lead to improvements in both i) the non-profit's scale of operations, as well as ii) the non-profit's operating efficiency and social performance. As such, financial subsidies and governance mechanisms might serve as complements toward the objective of increasing the non-profit's overall social impact.

That being said, it is unclear whether financial subsidies alone (i.e., without improving the non-profit's governance and management practices) are sufficient to help improve operating efficiency and the quality of services. Indeed, given the challenges and unique nature of non-profits, the sole provision of financial subsidies need not yield improvements along these dimensions as funding per se does not foster knowledge transfer and competence building, nor does it provide incentives to improve the organization's practices.

In the following, we explore the role of governance practices in improving non-profits' operating efficiency and their social impact in the context of a randomized governance program conducted in the DRC's healthcare sector (Sections 3-5). In auxiliary analyses, we also examine the interplay between governance and financial subsidies in the context of this program (Section 6).

3. Program and experimental design

In this section, we describe the PDSS program and experimental design. This section draws from the PDSS manual (PDSS 2016) that contains detailed information about the program.

3.1 The PDSS program

The PDSS program, known formally as "Projet de Développement du Système de Santé" (officially translated by the World Bank as "Health System Strengthening for Better Maternal and Child Health Results Project") is a five-year nationwide public health program that was launched by the government of

the DRC with the aim to improve the utilization and quality of maternal and infant healthcare services in targeted health zones in the country's territory. The program began in 2017 in 156 health districts in 11 of the country's 26 provinces: Kwango, Kwilu, Mai-Ndombe, Équateur, Mongala, Sud-Ubangi, Tshuapa, Haut-Katanga, Haut-Lomami, Lualaba, and Maniema. The program cost of $521 million is financed by the DRC government, the World Bank, UNICEF (United Nations International Children's Emergency Fund), UNDP (United Nations Development Program), Gavi, and donor country support.

The PDSS program attempts to improve the quantity and quality of primary healthcare services at the health facility level through two main mechanisms. First, the program presents health facilities with financial incentives to provide a "paquet minimum d'activités" ("minimum package of activities"). Defined by the World Health Organization (WHO), these packages put a strong emphasis on maternal and child health. Second, the program presents health facilities with a summary of their performance on a quarterly basis. This information is collected by the contracting and verification agencies ("établissements d'utilité publique," EUPs) established at the provincial level. These are semi-autonomous entities appointed by the DRC's Ministry of Health and Ministry of Finance to manage the contracting and procurement of health services with the health facilities. Concretely, they monitor and verify the health performance outcomes of each facility and provide feedback on the facility's quarterly management plans.

Prior to the PDSS program, other performance-based financing schemes were piloted in the DRC. These schemes presented a select number of health facilities with different rewards, initially taking the form of generic drugs only (2007-2009), thereafter generic drugs and cash (2009-2010), and finally cash only (2011-2017). Further, they were implemented initially in South Kivu only, later in North Kivu, Orientale, Kasai-Occidental and Kasai-Oriental, and lastly in Katanga. Notice that these initial pilots with performance-based financing were undertaken in only a limited number of health zones, which were different from the zones targeted by the PDSS program. Aside from these pilot schemes, we are not aware of major reforms in healthcare that coincided with our period of study, nor of other programs that may have differentially affected our treated health centers and thus biased our estimates.

Assessments of these pilot schemes were at best mixed. Mayaka et al. (2011) evaluated the first two pilot schemes using a qualitative research approach. They highlight the complexity of the schemes, a lack of shared understanding by key stakeholders, a lack of community engagement, and public sector mismanagement, especially corruption, as the main reasons for why the pilots led to unsatisfactory outcomes. Huillery and Seban (2021) examined the impact of the fee (cash only)-for-performance system on health service utilization and health providers' motivation in 152 health facilities in Haut Katanga. They find that, relative to the fixed payment system, the fee-for-performance mechanism slightly *reduced* service utilization, and argue that the main explanation for this result is health workers' lowered motivation at work due to a lack of understanding of how to best perform.

Guided by these lessons learned, the design of the PDSS program departed from previous performance-based financing schemes in healthcare in the DRC in three important ways. First, to address the lack of awareness and understanding about evidence-based practices, the PDSS program provides not only financial incentives but also auditing and feedback (A&F) to health facilities. Second, to mitigate problems of mismanagement, the control over the contracting and procurement relationships with each health facility was delegated by the government to third-party agencies at the provincial level. Furthermore, the PDSS program also introduced a community verification system, whereby local associations are mandated to check the actual existence of patients indicated in the facilities' registries and collect information about patients' satisfaction. Taken together, these three distinct design features aimed to improve the accountability of government and health workers, and the alignment of interests between the health system's key stakeholders: health workers, funders of the PDSS program, national and provincial governments, and patient communities.

The background context in which the PDSS program was conceived is one marked not just by poor health outcomes but also serious health system challenges. The DRC's human development indicators are among the lowest in the world, and four decades of conflict and mismanagement severely weakened the country's institutions and infrastructure. Moreover, the DRC is not on track to achieve any of the Millennium Development Goals, including those related to maternal and child health. The main maternal

and child health indicators remain very poor. According to UNICEF (2016), the maternal mortality ratio is 693 (per 100,000 live births) and the under-five child mortality rate is 98 (per 1,000 live births). Decreasing these mortality rates requires improvements in both the quantity and quality of reproductive and child health services. Together, these challenges help explain the aims and scope of the PDSS program.

3.2 Performance incentives

The PDSS program presents health facilities with financial incentives based on performance indicators measured at the health facility level (as opposed to the level of individual health workers). The size of the financial transfers (*T*) that health facilities receive depends on the performance along 22 health indicators and overall service quality. The transfers paid to the health facilities are mainly used for purchasing essential equipment. No more than fifty percent can be used toward health workers' salaries. Payments are made directly to the health facilities.[12]

The overall transfer is comprised of two parts:

$$T = \sum_{i=1}^{22} p_{is} q_i + \alpha$$

where the first part of the transfer reflects the performance along the 22 healthcare performance indicators listed in Table A1. For each indicator i, and the corresponding number of health services q_i, a target is set depending on the size of the population served by the health facility. The payment amount per unit of health service provided, p_{is}, varies depending on the percentage share s of the target performance achieved. For each indicator i, the closer the actual performance is to the targeted performance, the higher the unit price. Some of these indicators refer to the reasons for a visit (such as prenatal care or delivery), whereas others refer to the type of services provided during a visit (such as tetanus vaccination during prenatal care). All the indicators and payment amounts are defined in concordance with WHO guidelines.

[12] Note that, even if some of the transfers are used towards the health center's wages, this component remains distinct from traditional financial incentives. This is because the subsidies are based on the health centers' social performance and are paid directly to the health centers (as opposed to being paid to individual employees based on their individual performance).

The second part of the transfer reflects the overall health service quality. The parameter α denotes the "quality bonus," a supplement of up to 25% of $\sum_{i=1}^{22} p_{is}q_i$. Quality is assessed using a quality assessment grid that consists of a longlist of evidence-based practices. The grid considers practices in 15 distinct domains, ranging from the overall organization and management of the health facility to practices in terms of HIV/tuberculosis, and hygiene and sterilization. The list is provided in Tables A2 and A3. A total of 400 points can be earned. To receive a bonus, the health center must secure a minimum of 200 points, that is, obtain a quality score of at least 50 percent. When the quality score is above 50 but below 80 percent, the quality bonus is given by α = quality score × 25% × $\sum_{i=1}^{22} p_{is}q_i$. When the quality score is above 80 percent, then the quality bonus is given by α = 25% × $\sum_{i=1}^{22} p_{is}q_i$.

Health facilities submit monthly activity reports and quarterly requests for payment to the EUPs, who are responsible for the verification of data and authorization of payments. A team of auditors verifies the reported number of medical acts in the minimum package of activities delivered by the center during the period, and reviews the quality of the services provided using the quality assessment grid. The auditors are recruited by and are accountable to the EUPs. They are helped by local associations whose role is to question the community to i) find out whether the patients declared by the centers have effectively received the corresponding treatments, and ii) gather feedback regarding their satisfaction. The team of auditors not only assesses the health facilities' performance outcomes and practices, it also provides feedback and recommendations. The grid provides objective data regarding discrepancies between current practices and target performance. Every quarter, the health facilities are expected to draw on these inputs to plan realistic and progressive improvements. The auditors support the write-up of the quarterly management plans that define how the improvements can be achieved. They also provide training in finances and stock management.[13]

[13] The auditors are members of the World Bank's Development Policy Staff who were previously deployed in the DRC and have expertise in healthcare. On top of their medical training, they received additional training by World Bank specialists in performance-based financing. Note that it is usually the same team of auditors who visit a given health center each quarter. This is meant to ensure a certain continuity in the feedback that is provided. A given team of auditors is often affected to several health zones within a given province.

3.3 Selection into the program and randomization

The PDSS program was implemented in 11 of the DRC's 26 provinces. The 11 provinces were selected based on three criteria: poor health indicators, limited access to health services, and the ability to build on or expand an ongoing partnership with an international organization. Only health centers registered with the DRC Ministry of Health were eligible for the program. (They need not be affiliated with the government, though.) Within each province, the selection of health facilities was made by Médecins d'Afrique, an international NGO, in close collaboration with the World Bank's impact assessment team and the PDSS project unit. Together, they conducted a baseline evaluation for which they designed questionnaires and protocols for establishing the selection. They then assessed the health centers' suitability for the program as health centers were required to have a minimum level of quantity and quality of services. The baseline evaluation was conducted between June 2015 and March 2016. The selected health centers were then randomly assigned into a treatment group ("Groupe cas") and control group ("Groupe témoin"). The randomization itself was done in Excel, and every third health center was assigned into the control group.

It is important to note that both the treated and control health centers are part of the PDSS program, in that they both receive funding from the program. The payments are twofold. First, at the beginning of the program (first quarter of 2017), health centers in both the treatment and control groups receive an initial subsidy (called "unité d'investissement") that is meant to finance the purchase of essential equipment. Second, in the first month of each quarter (starting in the second quarter of 2017), the performance-based subsidy is paid out to the treated health centers based on the auditors' assessment conducted in the previous quarter (the formula used to compute the subsidy is described above in Section 3.2). Control health centers also receive a quarterly subsidy that matches the average subsidy received by the treated health centers in the same quarter and the same health district. As such, the quarterly payments are on average the same in both the treatment and control groups.

By design, the PDSS program provides an ideal setup for our study. By comparing health centers that are randomly assigned to the treatment versus control group, we are able to identify the impact of the

governance intervention—that is the provision of social performance-based incentives combined with A&F—on health centers' operating efficiency and social performance, holding everything else, *including funding*, constant. Table 1 provides a summary of the PDSS intervention and what it entails for both the treated and control health centers.

----- Insert Table 1 about here ----

4. Data and methodology

4.1 Data sources

The data on the DRC health centers are obtained from administrative records of the DRC Ministry of Health that are maintained in the DRC's Système National d'Information Sanitaire (SNIS). They include detailed information on the health centers' operations (e.g., staff, number of consultations, number of births), along with the name and location of each health center.

These data were supplemented with identifiers for the treated and control health centers provided by the World Bank. Our baseline sample includes a total of 999 health centers, out of which 674 were assigned to the treatment group, and 325 to the control group. For each health center, we were granted access to 14 quarters of data, ranging from the first quarter of 2016 until the second quarter of 2019. That is, our dataset includes four quarters before and ten quarters after the program's launch.

Note that, while the large-scale implementation of the PDSS program took place in the first quarter of 2017, a small number of health centers were already treated in 2016 in a pilot-like setting. Those are excluded from our sample.

4.2 Outcome variables

Our objective is to study how the governance intervention—that is, the provision of performance-based incentives combined with A&F—affects health centers' outcomes. In what follows, we describe the outcome variables.

Health center's operating efficiency

We compute a health center's operating efficiency as the number of primary healthcare services performed divided by the number of employees. This measure captures the health center's labor productivity (i.e., output per employee). Note that the health centers in our sample only offer primary healthcare services. Secondary healthcare services are typically administered at hospitals, often upon referrals from the health centers.

Health center's employees

We use several variables to examine changes in the health center's staff. First, we use the total number of employees working at the health center. Second, we decompose this total into the number of doctors, nurses, and administrative personnel, respectively.

Volume of healthcare services

To measure the volume (or "quantity") of healthcare services, we use the number of primary healthcare services performed. Since antenatal care and childbirth are the main services performed at the health centers, we also use two additional metrics: the number of maternity and childhood healthcare services performed, and the number of births.

Quality of healthcare services

To measure the quality of healthcare services, we focus on infant mortality at birth. This is a key metric in our context since antenatal care and childbirth are the main services performed at the health centers. Infant mortality at birth can occur in two forms: stillbirth and neonatal death. Stillbirth refers to a baby born with no sign of life at or after 28 weeks of gestation; neonatal death refers to a baby who dies within the first 28 days of life (WHO 2019a, 2019b). We compute the ratio of stillborn babies to the total number of births (henceforth "share of stillbirths"), the ratio of neonatal deaths to the total number of births ("share of neonatal deaths"), and the complement ("share of live births"). These three ratios allow us to assess not only the quality of the childbirth services per se, but also the quality of antenatal care services. Indeed,

medical research has shown that antenatal care reduces the likelihood of stillbirth and neonatal death (Adam et al. 2005, Hollowell et al. 2011). As such, the above measures capture the quality of the main services performed at the health centers.

4.3 Summary statistics

Table 2 provides summary statistics for the 999 health centers in our sample. All statistics refer to the fourth quarter of 2016 (i.e., the quarter that precedes the start of the PDSS program). On average, health centers in our sample performed 1,611 primary healthcare services (236 on a per employee basis). The majority (957 out of 1,611) were maternity and childhood healthcare services.[14] As discussed above, antenatal care and childbirth are the main services performed at the health centers; the summary statistics reflect this institutional feature of the DRC's healthcare system.

----- Insert Table 2 about here ----

Other statistics are worth highlighting. The average number of employees is 7.1, consisting mainly of nurses and administrative personnel. The average (quarterly) number of births is 53.8, out of which 0.70% are stillbirths, and 0.17% neonatal deaths. Lastly, as can be seen from the bottom panel, the average health center is located in a health district of 204,409 inhabitants, and a health area of 11,135 inhabitants.[15] To further characterize the health centers from our sample, Figure 1 provides photographs featuring two of them.

----- Insert Figure 1 about here ----

4.4 Randomization tests

Our identification strategy relies on the random assignment of health centers to the treatment and control groups. Since randomization is a feature of the PDSS program, this requirement should hold by design. To empirically assess the validity of the randomization, we can examine the covariate balance prior to the

[14] Note that the number of healthcare services is not equal to the number of patients, as the same patient can receive multiple services within the same consultation.

[15] The DRC is partitioned into 516 health districts (also called health "zones") and 8,504 health areas. Appendix A provides a description of the DRC's health system along with a characterization of the health districts and health areas, respectively.

treatment—intuitively, if the assignment is truly random, there should be no systematic difference between health centers in the treatment and control groups based on pre-treatment characteristics.

We conduct this analysis in Table 3, where we report the same set of summary statistics as in Table 2, but separately for the 674 health centers in the treatment group and the 325 health centers in the control group. The statistics are again computed in the fourth quarter of 2016 (i.e., the quarter that precedes the launch of the PDSS program). The last two columns of the table provide the *p*-value of the difference-in-means test and the Kolmogorov-Smirnov (KS) test of identical distributions, respectively, for each covariate.

----- Insert Table 3 about here ----

As can be seen, there is no systematic difference between the treated and control health centers. For all covariates, the summary statistics are very similar in economic terms. They are similar in statistical terms as well. Specifically, the difference-in-means test (reported in the penultimate column) is always insignificant, with *p*-values ranging from 0.196 to 0.981. Similarly, the KS test of identical distributions (reported in the last column) is always insignificant as well, with *p*-values ranging from 0.140 to 0.980. Overall, the evidence from Table 3 confirms the random assignment of health centers.

This randomization is further illustrated in Figure A2, where we plot the distribution of each outcome variable in the treatment vs. control groups in the quarter preceding the treatment. As can be seen, the distributions are very similar for each variable. The formal test for identical distributions—the KS test from Table 2 mentioned above—confirms that the null of identical distributions cannot be rejected for all variables.

Finally, in Figure 2, we plot the location of the health centers in the control group (blue markers) and treatment group (green markers) on the DRC map. We caution that the map is incomplete as granular geo-codes (and hence the longitude-latitude coordinates) are only available for about 73% of the sample. As can be seen, there is no apparent imbalance between the two groups. More formally, when we examine the distribution of the control and treated health centers within each province, we find no significant deviation from the program's targeted 2/3 of treated health centers. Specifically, the null of a 2/3 distribution

within each province cannot be rejected with a *p*-value of 0.268.

----- Insert Figure 2 about here ----

4.5 Methodology

To examine how the treatment affects health center outcomes, we use a difference-in-differences methodology. Specifically, we estimate the following regression:

$$y_{it} = \alpha_i + \alpha_t + \beta \times treatment_i \times post_{t = 0 \text{ to } 4} + \gamma \times treatment_i \times post_{t = 5 \text{ to } 9} + \varepsilon_{it} \quad (1)$$

where y is the outcome variable for health center i in quarter t. The outcome variables are either ratios (e.g., share of stillbirths) or levels (e.g., number of employees). We take the natural logarithm of all level variables. The quarters range from $t = -4$ (Q1 2016) to $t = 9$ (Q2 2019), that is, the sample includes four quarters before and ten quarters after the intervention ($t = 0$ is the quarter of the intervention, Q1 2017). On the right-hand side of the equation, α_i are health center fixed effects; α_t are quarter fixed effects; $treatment_i$ is an indicator variable equal to one if the health center is in the treatment group (and zero if it is in the control group); $post_{t = 0 \text{ to } 4}$ and $post_{t = 5 \text{ to } 9}$ are indicator variables equal to one for the first five quarters (t = 0 to 4) and the subsequent five quarters (t = 5 to 9), respectively, following the intervention; and ε_{it} is the error term. We block-bootstrap standard errors (using 1,000 bootstrap replications) at the health center level to account for potential dependence of the error term along the dimension of the treatment. The coefficients of interest are those of $treatment_i \times post_{t = 0 \text{ to } 4}$ and $treatment_i \times post_{t = 5 \text{ to } 9}$ that capture the treatment effect in the first five and subsequent five quarters of the intervention, respectively.

In our main analyses, we also estimate a variant of equation (1) that provides a finer-grained characterization of the dynamics of the treatment effect:

$$y_{it} = \alpha_i + \alpha_t + \sum_{t \neq -1} \beta_t \times treatment_i \times \alpha_t + \varepsilon_{it} \quad (2)$$

where $\beta_{t = -4}$, $\beta_{t = -3}$, $\beta_{t = -2}$ allow us to assess potential pre-trends and $\beta_{t = 0}$, $\beta_{t = 1}$, …, and $\beta_{t = 9}$ provide a characterization of the treatment dynamics on a quarterly basis (relative to $t = -1$).

5. Results

5.1 Impact of the governance treatment on health center outcomes

Baseline estimates

Table 4 presents our main results. The estimates are obtained from regression (1) using the 999 health centers in our sample over the 14 quarters for which we have data (that is, the number of observations is 999 × 14 = 13,986).

----- Insert Table 4 about here ----

As can be seen, the benefits from the governance treatment are not immediate. In all columns, the treatment effect is small and insignificant in the first five quarters (t = 0 to 4). It is only in the subsequent five quarters (t = 5 to 9) that we observe significant changes in health center outcomes.

In column (1), the dependent variable is operating efficiency. As is shown, the governance treatment brings about large improvements in quarters t = 5 to 9. The point estimate of 0.255 (p-value = 0.000) implies that the number of primary healthcare services per employee increases by 25.5%.

In columns (2)-(8), we unpack this improvement in operating efficiency by examining the numerator (number of primary healthcare services performed) and denominator (number of employees). We find that the treated health centers are able to expand their volume of healthcare services without a commensurate increase in employment. Specifically, we find in column (6) that the number of primary healthcare services performed increases by 23.3% (p-value = 0.000). This increase in the volume of healthcare services is further confirmed by the estimates in columns (7) and (8) showing that the number of maternal and childhood healthcare services and the number of births, respectively, increase by 14.9% (p-value = 0.052) and 13.8% (p-value = 0.004).

While the volume of healthcare services increases, the number of employees does not. This can be seen from column (2), in which we observe no significant change in employment. In columns (3)-(5), where we decompose employment by job types, we find no significant change in the number of doctors and nurses. Interestingly, the treated health centers have less of a need to hire administrative employees—relative to

the control group, their number of administrative employees decreases by 5.8% (p-value = 0.043). Note that the decrease in administrative employees is not about layoffs. In fact, both the treatment and control group increase their administrative staff during the sample period. The observed difference between the two groups indicates that the treated health centers hire fewer additional administrative employees compared to the control group.

Overall, the findings from columns (1)-(8) indicate that the treated health centers become more efficient following the governance intervention. They are able to offer a higher volume of services without a commensurate increase in employment. In addition, they have less of a need to hire administrative employees in order to expand their volume of services.

In columns (9)-(11), we turn to the quality of healthcare services. We find evidence for a significant increase in service quality following the treatment. Specifically, we find that, in quarters t = 5 to 9, the share of stillbirths decreases by 0.21 percentage points (p-value = 0.003), and the share of neonatal deaths by 0.11 percentage points (p-value = 0.044). Correspondingly, the share of life births increases by 0.32 percentage points (p-value = 0.001). Put differently, for every 1,000 new births, the governance treatment helped save about 2.1 lives at birth (reduction in stillbirths) and 1.1 lives within the first 28 days after birth (reduction in neonatal deaths). These are large effects in light of the baseline probabilities. Indeed, the pre-treatment share of stillbirths and neonatal deaths is on average 0.91 percentage points (Table 2). Hence, a decrease by 0.32 percentage points in the share of stillbirths and neonatal deaths corresponds to a 35.2% reduction in infant mortality. Overall, these results indicate that the governance intervention brought about large improvements in health outcomes.

In Table A4, we further report the minimum detectable effect sizes (MDES) pertaining to a 5% significance level and 80% power. As is shown, our main estimates—that is, the increase in operating efficiency and the increase in the share of live births in quarters t = 5 to 9—are well above their respective

MDES.[16]

Dynamics

In Table 5, we provide a finer-grained characterization of the treatment dynamics. To do so, we estimate regression (2), in which we estimate the treatment effect on a quarterly basis (relative to $t = -1$). This finer-grained analysis also allows us to assess potential pre-trends.[17]

As can be seen, we find no evidence for pre-trends; the corresponding point estimates are all small and insignificant. Moreover—and consistent with our findings from Table 4—we find no evidence for an immediate response at the time of the intervention. Rather, the benefits of the treatment take a few quarters to materialize. The first tangible effect is observed after four quarters, when the increase in operating efficiency becomes significant (along with the increase in the number of primary healthcare services and the lower need for administrative employees). The improvements in the quality of healthcare services take even longer to materialize. It is only after seven quarters that the birth mortality statistics start showing significant improvements. These dynamic patterns are confirmed in Figure 3, where we plot the evolution of each outcome variable in the treatment and control groups.[18]

----- Insert Table 5 and Figure 3 about here ----

The delayed response observed in Table 5 and Figure 3 is consistent with the gradual adoption and learning of better practices. This was confirmed in a series of interviews we conducted with program participants ($N = 20$).[19] Collectively, the respondents highlighted that a key challenge was the lack of adequate training and organizational know-how. In this regard, the quarterly rounds of feedback (combined

[16] In Appendix B, we present several robustness checks. Specifically, we show that our results are robust if we i) use clustered standard errors (Table A5), ii) use Anderson's (2008) sharpened *q*-values that account for multiple hypothesis testing (Table A6), and iii) re-estimate the regressions of the share of stillbirths, neonatal deaths, and live births using weighted least squares (WLS), weighting observations by the number of births, thereby accounting for the fact that ratios of births are more accurately measured when the number of births is higher (Table A7).
[17] In Figure A3, we plot these estimates in event time (along with 95% confidence bounds) for each outcome variable.
[18] For ease of exposition, in each plot, the level of the control group is normalized to match the one of the treatment group at $t = -1$.
[19] We conducted 20 interviews with individuals involved in the PDSS program, out of which 17 were based in the DRC. The interviewees included health center employees (nurses and doctors), auditors, and programs administrators. The recordings of these interviews are available upon request.

with incentives to actually act on this feedback) were seen as essential in inducing tangible changes that would ultimately translate into higher efficiency and quality of services.

It is also worth noting that the long learning curve of the health centers' employees was a key consideration in designing the PDSS program. As discussed in Section 3.1, prior to the PDSS program, pilot programs using performance-based financing designs were run in selected parts of the DRC. These programs were unsuccessful, suggesting that the mere use of performance-based financing is insufficient to induce tangible improvements. This led to the innovative design of the PDSS that combines performance-based incentives with A&F.

Lastly, the longer lag we observe for the improvements in infant mortality (7 quarters, compared to 4 quarters for the improvements in operating efficiency) reflects the type of services provided at the health centers. Indeed, the bulk of the health centers' services pertain to pre-natal care. By their very nature, improvements in pre-natal care reduce the risk of stillbirths and neonatal deaths several months in the future, and are likely more effective when higher-quality services are provided throughout the full pregnancy cycle. In other words, the longer lag we observe is likely explained by the fact that it first takes about 4 quarters for the governance improvements to materialize, and then another 9 months (i.e., another 3 quarters, adding up to 7 quarters) for patients to fully benefit from the improved pre-natal care throughout their pregnancy and until carrying the child to term.

Economic magnitudes

As mentioned above, our baseline estimates are large in economic terms—the estimates from Table 4 imply that operating efficiency increases by 25.5% and infant mortality decreases by 35.2%. Our interviews of program participants help shed light on these magnitudes. All respondents (N = 20) expected the improvements to be large, citing the low education level and inadequate training of the health center employees as key rationales for a steep learning curve. One respondent even qualified the benefits of the program as "énormes" (enormous) due to the above reasons.

While interviews are subjective in nature, a perhaps more objective benchmark is provided by the

World Bank, who targeted to reach a 65% quality score for the treated health centers, compared to a 20% score prior to the start of the PDSS program (World Bank 2018). This corresponds to more than a three-fold increase in quality.

To further assess the magnitudes, and put them into perspective, we benchmark our estimates against related estimates from the literature. In this regard, the literature on management practices is especially informative, as considerable effort has been put into quantifying the productivity gains from different management practices. (In spirit, our governance treatment consisting of social performance-based incentives combined with A&F can be seen as a form of "structured management practice" in the terminology of Bloom et al. 2019). Specifically, in their study of management practices at U.S. manufacturing plants, Bloom et al. (2019) find that a one standard deviation increase in their score of structured management is associated with a 26% percent increase in labor productivity, noting that "[t]he magnitude of the productivity-management relationship is large" (p. 1649). Giorcelli (2019) finds that Italian companies who participated in a management assistance program achieved a 49% increase in productivity. Similarly, Bruhn et al. (2017) find that Mexican small and medium enterprises that were offered a one-year access to management consulting services achieved a 27% increase in productivity.

Naturally, we caution that the benefits of the improved management practices considered in these articles need not be directly comparable to those of the intervention considered in our study given the different nature of non-profit organizations. Nevertheless, they do provide guidance in terms of how changes in business practices can affect operating efficiency. Collectively, they indicate that the efficiency gains can be substantial, in the ballpark of what we find in our setting. In addition, such efficiency gains might be larger in economically developing countries such as the DRC due to the potentially bigger room for improvement, as several respondents pointed out in our interviews.[20]

[20] Another way to put the magnitudes into perspective is through the VSL (value of a statistical life) framework. Viscusi and Masterman (2017) estimate a VSL of $71,000 for the DRC (compared to a VSL of $9.6M for the U.S.). Our baseline results in Table 4 suggest that the governance treatment reduces infant mortality by 0.32%. Factoring in the number of treated health centers (674) and the average number of births per health center (55.7, see Table 3), the governance intervention corresponds to VSL gains in the amount of $8.5M. Compared to the cost of the program

Opening up the black box of the treated health centers: what explains the improvements in operating efficiency and health outcomes?

Our results indicate that the PDSS program led to significant improvements in operating efficiency and health outcomes at the treated health centers. However, a limitation of our analysis is that we do not directly observe which operational changes are made at the treated health centers.

To gain insights into what happened within the "black box" of the treated health centers, we obtained additional data from the DRC's Ministry of Health. Specifically, we were given access to the 15 quality scores that were collected at the treated health centers as part of the PDSS program. In Table 6 and Figure A4, we examine the change in these quality scores over the 10 quarters for which we have data (that is, from $t = 0$ until $t = 9$). Naturally, we caution that this analysis is merely descriptive, as these data were only collected at the treated health centers (i.e., no data are available for the control group) during the PDSS program (i.e., there is no pre-treatment data).[21]

----- Insert Table 6 about here ----

As can be seen from Table 6, we observe large increases in the quality of planning, financial management, and general organization. Over the 10 quarters, the corresponding quality indices (which range from 0 to 100) increased by 72.6 (p-value = 0.000), 55.0 (p-value = 0.000), and 50.7 (p-value = 0.000) index points, respectively. This suggests that improvements in planning, financial management, and the health centers' general organization are plausible explanations for the observed improvements in operating efficiency.

The other quality scores in Table 6 pertain to the clinics' medical operations. While they all increase following the intervention, the largest improvements are found in categories that are of direct relevance to maternal and childhood health. Those include maternity care, pre-natal care, and family planning (those

($521M), this translates into a VSL-ROI of about 1.6% from governance alone. Naturally, we caution that this number likely represents a lower bound since the improvements brought about by the governance intervention are likely to improve patient health in ways that are not captured by our infant mortality metrics.

[21] A description of these quality scores is provided in Table A3. Note that these data are available for 646 of the 674 treated health centers. Also, since the category "tracer drugs" has missing data for $t = 0$, 1, and 2, it is not included in this analysis.

categories show quality improvements of 41.2, 39.4, and 38.3 index points, respectively, all with *p*-value = 0.000). These improvements are plausible channels that explain the decrease in the number of stillbirths and neonatal deaths documented earlier.[22]

The dynamic analysis in Figure A4 further shows that these improvements tend to materialize within the initial quarters of the PDSS program and remain somewhat stable thereafter. This is consistent with the "learning" interpretation discussed earlier, in which the initial quarters of the program are essentially a learning phase that subsequently translates into higher operating efficiency and lower infant mortality.

Cross-sectional heterogeneity

In Appendix C (and Tables A8 and A9), we examine how our baseline estimates vary depending on several cross-sectional characteristics. We find that the increase in operating efficiency and the decrease in infant mortality are significantly stronger in districts with a lower density of health centers (measured by the number of health centers divided by the health district's population). These findings are consistent with the learning interpretation, as they suggest that health centers benefit more from the treatment when it is harder to learn and adopt best practices from their peers.

In addition, we find that i) the increase in operating efficiency is higher for health centers with lower ex ante operating efficiency, and ii) the decrease in infant mortality is more pronounced for health centers with higher ex ante mortality. These findings indicate that the improvements brought about by the treatment are larger when there is more to improve and learn to begin with. This again points toward the learning interpretation discussed above.

[22] Large increases are also found in two other categories, namely indigent committee and drugs and consumables. The former refers to the attention given to the most vulnerable groups in an effort to foster more inclusive healthcare. Arguably, both dimensions are likely to further contribute to lower infant mortality.

5.2 Potential challenges and alternative interpretations

Contamination

An important concern is that the treatment might be contaminated by strategic interactions among nearby health facilities. In particular, health centers may try to "game" the performance-based incentives by strategically refusing high-risk patients (e.g., pregnant mothers who face a high risk of stillbirth), referring them to nearby health centers and hospitals. Such gaming behavior could potentially explain our results.

Nevertheless, this concern is mitigated, for two reasons. First, as part of their assessment protocol, the auditors conduct interviews among the local community. Accordingly, if health centers were to transfer high-risk patients elsewhere, they would likely find out. Second, in Table 7, we examine whether the share of stillbirths and neonatal deaths increases (and the share of live births decreases, respectively) at other healthcare facilities that are located in the same health district as the treated health centers. Those include hospitals (columns (1)-(3)), control health centers (columns (4)-(6)), and health centers not in the PDSS program (columns (7)-(9)) that are located in the same health district as the treated health centers. As can be seen, we find that all point estimates are small and insignificant, which is inconsistent with the gaming interpretation.[23]

----- Insert Table 7 about here ----

Other challenges

In Appendix D, we discuss other potential challenges of the PDSS experiment. Specifically, we discuss the possibility that i) corruption, ii) employee turnover, iii) implementation challenges (e.g., limited accessibility to certain health centers), iv) the Hawthorne effect, and v) the potential rise of "superstar" health centers might affect our results. As we discuss in Appendix D, these challenges are unlikely to

[23] A related concern is that the treated health centers may expand their capacity post-treatment, and hence be able to see a higher number of low-risk patients who would have otherwise relied on home care. In this scenario, the treated health centers would have a higher share of low-risk patients, which could explain the improvements in the birth statistics. Nevertheless, in our interviews, it was noted that the choice to deliver at home or at the health center is not made based on the risk assessment of the pregnancy, as women themselves cannot easily assess the level of risk of their pregnancy. Rather, the stated rationales were the perceived quality of services offered by the health centers and the value for money of these services. Neither is related to the pregnancy's riskiness.

materially affect our results.

6. Funding vs. governance: auxiliary evidence from health centers outside the PDSS program

In the analysis presented so far, we focused on health centers that were selected for the PDSS program—that is, health centers that were either in the control group (receiving PDSS funding) or in the treatment group (receiving PDSS funding and the governance treatment).

In this section, we consider a third group of health centers, namely those that are not part of the PDSS program. We refer to this group as the "outside group." By construction, the outside group does not receive any funding nor the governance intervention from the PDSS program. Accordingly, we can use the outside group to examine the relationship between funding and governance. Specifically, by comparing the treatment group (that receives funding and the governance treatment) versus the outside group (that receives neither), we can assess the benefits of the 'funding and governance' bundle. Similarly, by comparing the control group (that receives funding) versus the outside group, we can assess the effectiveness of 'funding only.'

Our dataset from the DRC Ministry of Health covers a total of 5,832 health centers in the outside group. Table A10 provides summary statistics for those. Compared to the 999 health centers in our sample, the "outside" health centers are on average smaller (5.2 versus 7.1 employees) and provide a lower volume of services (1,153 versus 1,611 primary healthcare services performed). These differences are not surprising. Indeed, as described in Section 3.3, inclusion in the PDSS program is not random. (What is random is the assignment to the treatment versus control group *within* the PDSS program.) Hence, a caveat of using the outside group is that we can no longer rely on randomization, and hence any such analysis is correlational per se.

To mitigate this caveat, we use a nearest-neighbor matching in which health centers in the treatment group (and control group, respectively) are matched to health centers in the outside group based on a large set of observables. Specifically, for each of the 674 health centers in the treatment group (and for each of the 325 health centers in the control group, respectively), we match the nearest—i.e., most similar—health

center out of the pool of 5,832 health centers in the outside group. The matching is done in two steps. First, we require that the matched health center be located in the same health district as the treated health center. (If less than 10 outside health centers are available within the relevant district, we relax this requirement and require that the matched health center be located in the same province as the treated health center.) Second, out of the remaining candidates, we select the nearest neighbor based on the health center characteristics in Table A10 (measured in the quarter that precedes the PDSS program). The nearest neighbor is then the one with the lowest Mahalanobis distance to the treated health center along the matching characteristics.[24]

This matching procedure ensures that the matched health centers from the outside group are as similar as possible to the treated health centers (and control health centers, respectively) ex ante. The covariate balance, provided in Table A11, confirms that there is no significant difference between the respective groups.

To compare outcomes in the i) treatment versus outside groups and ii) control versus outside groups, we estimate a variant of regression (1) in which we pool the two matched samples together and replace the treatment indicator with a *treatment vs. outside* indicator (which is equal to one for health centers in the treatment group and zero otherwise) and a *control vs. outside* indicator (which is equal to one for health centers in the control group and zero otherwise). To ensure that comparisons are made within each matched pair (that is, treated and control health centers are compared to their respective matched outside health centers), we include matched pairs by quarter fixed effects in all regressions. The results are provided in Table 8.

----- Insert Table 8 about here ----

Control group versus outside group: assessing the effectiveness of funding alone

The coefficients pertaining to the *control vs. outside* indicator provide a comparison of health centers in the

[24] Formally, the Mahalanobis distance δ between treated health center i and candidate health center j is given by $\delta = [(\mathbf{X}_i - \mathbf{X}_j)' \, \mathbf{\Sigma}^{-1} \, (\mathbf{X}_i - \mathbf{X}_j)]^{½}$, where $\mathbf{X}$ is the vector of matching characteristics and $\mathbf{\Sigma}$ the covariance matrix.

control group versus health centers in the outside group. Since the control group receives PDSS funding, and the outside group does not, this comparison allows us to examine how funding as standalone (i.e., without governance improvements) affects health center outcomes.

As is shown, we find that funding as standalone leads to increases in the scale of the health center's operations—the number of employees increases by 17.8% (p-value = 0.000), the number of primary healthcare services by 38.7% (p-value = 0.046), the number of maternal and childhood healthcare services by 47.0% (p-value = 0.032), and the number of births by 21.9% (p-value = 0.043)—but does not bring about significant improvements in operating efficiency (column (1)) nor in the quality of healthcare services (columns (9)-(11)). This indicates that funding by itself, while helpful in increasing the health center's scale and volume of services, is not sufficient to induce improvements in operating efficiency nor the quality of service.

Treatment group versus outside group: assessing the effectiveness of funding and governance combined

Analogously, the coefficients pertaining to the *treatment vs. outside* indicator provide a comparison of health centers in the treatment group versus health centers in the outside group. This comparison allows us to examine how the combination of both funding and governance affects health center outcomes.

We find again that, compared to the outside group, the treated health centers substantially increase the scale of their operations—the number of employees increases by 15.0% (p-value = 0.001), the number of primary healthcare services by 60.1% (p-value = 0.000), the number of maternal and childhood healthcare services by 56.7% (p-value = 0.001), and the number of births by 34.0% (p-value = 0.000). In addition, and similar to what we observed in Table 4, operating efficiency (column (1)) and the quality of healthcare services (columns (9)-(11)) significantly improve following the treatment. Overall, these results indicate that the 'funding and governance' bundle benefits the treated health centers in two ways: the additional funding allows them to grow their operations, while the governance intervention helps them improve their economic and social performance.

Taken together, the results from Tables 4 and 8 suggest that i) funding alone increases the scale of

the health centers but does not improve operating efficiency nor service quality (Table 8); ii) governance alone improves operating efficiency and service quality (Table 4); and iii) the combination of funding and governance leads to both an increase in the scale of the health centers and improvements in the health centers' operating efficiency and service quality (Table 8). As such, these results suggest that funding is not a substitute for governance. Instead, they complement each other in improving the health centers' social impact.

7. Discussion and conclusion

How can non-profit organizations improve their governance in order to maximize their intended social and environmental impact? This question is of foremost importance not only for the non-profits themselves (as well as their donors and impact investors), but also from an economic, environmental, and social perspective.

In this study, we shed light on this question by exploiting a randomized governance program conducted in the DRC's healthcare sector. The program was administered in about 1,000 non-profit health centers that were randomly assigned into a treatment and control group. While health centers in both groups received financial subsidies from the program, only those in the treatment group were subject to the "governance treatment" consisting of social performance-based incentives combined with A&F. As such, this randomized program provides an ideal setup to study how the adoption of governance mechanisms affects health centers' outcomes, holding everything else (including financial subsidies) constant.

We find that the governance treatment led to i) a significant increase in the number of services performed per employee, and ii) a significant reduction in the share of stillbirths and neonatal deaths. These findings indicate that governance plays an important role in improving non-profits' operating efficiency and social performance, respectively.

In auxiliary analyses, we further examine whether financial subsidies could potentially serve as a substitute for governance. We find that this is unlikely to be the case. Specifically, we find that health centers that only receive financial subsidies increase their scale, but do not show improvements in operating

efficiency nor in the quality of their services. In contrast, health centers that receive both financial subsidies and the governance treatment improve both their scale as well as their operating efficiency and the quality of their services. This suggests that financial subsidies and governance operate as complements toward the objective of increasing the non-profit's overall social impact.

Our study contributes to several strands of the literature. First, by exploring the role of governance (in the form of social performance-based incentives and A&F) in enhancing the social impact of non-profits operating in a resource-constrained environment, this study relates to the large literature that has studied the role of governance and management practices in the context of for-profit organizations (e.g., Aguilera et al. 2016, Blader, Gartenberg, and Prat 2020, Bloom and Van Reenen 2007, Bloom et al. 2013, Bryan, Tilcsik, and Zhu 2017, Cai and Szeidl 2018, Chatterji et al. 2019, De Mel, McKenzie, and Woodruff 2014, Dimitriadis and Koning 2022, McKenzie 2021, McKenzie and Woodruff 2014, 2017). Since the non-profit and for-profit contexts are fundamentally different, it is a priori unclear i) which governance and management practices are available and effective for non-profits in a resource-constrained setting, as well as ii) when and where those practices matter more or less. Our study sheds light on these questions and highlights the effectiveness of a bundle of governance mechanisms that consists of social performance-based incentives combined with A&F. Moreover, we show that, although financial subsidies cannot substitute for governance, they can enhance the effectiveness of the governance bundle.

In addition, our study contributes to the literature that examines the social performance of i) for-profit organizations (e.g., Eccles, Ioannou, and Serafeim 2014, Flammer 2015, Flammer, Hong, and Minor 2019, King and Lenox 2001); ii) for-profit organizations in collaboration with non-profits and non-governmental organizations (e.g., Ballesteros and Gatignon 2019, Cabral et al. 2019, Chatain and Plaksenkova 2019, Durand and Huysentruyt 2022, Rousseau, Berrone, and Gelabert 2019); iii) hybrid organizations (e.g., Battilana and Dorado 2010, Cobb, Wry, and Zhao 2016, Jay 2013, Pache and Santos 2013, Quélin, Kivleniece, and Lazzarini 2017); iv) non-profit organizations in the healthcare sector (e.g., Kellogg 2009); v) non-profit organizations in developing countries (e.g., Mair, Marti, and Ventresca 2012); as well as vi) the literature on non-profit governance (e.g., Bertrand et al. 2020, Fisman and Hubbard 2005,

Krause et al. 2019). Our study complements this vibrant line of work by exploring how the governance of non-profits (along with the interplay between governance and financial subsidies) can contribute to their social impact in a resource-constrained environment and the attainment of the United Nations' Sustainable Development Goals (SDGs). In this regard, our study also adds to the growing literature that examines how management research can help understand and address grand societal challenges (e.g., Berrone et al. 2016, George et al. 2016a, Vakili and McGahan 2016).

Moreover, our study examines an underexplored institutional context: Africa, and more specifically the DRC in Sub-Saharan Africa. While most of the insights from the governance literature have been obtained by studying organizations in developed and emerging countries, much less is known about (for-profit and non-profit) organizations in developing countries, and especially Africa (George et al. 2016b).[25] Yet, understanding how to improve the governance of these organizations—and the implications for operating efficiency and social performance—is crucial in order to promote their development and the attainment of the United Nations' SDGs.

Our study calls for future research. First, our findings are specific to the healthcare sector in a low-income country (namely, the DRC). In this regard, a fruitful avenue for future research is to examine whether our findings have external validity across sectors and countries, including higher-income countries. Arguably, the challenges faced by non-profit organizations in higher-income countries are similar but likely less severe than in low-income countries. As such, the effectiveness of the governance bundle considered in this study might differ. Future research may find it worthwhile to explore and characterize these differences. Second, a limitation of the PDSS program is that the "governance treatment" is administered through a bundle of governance mechanisms (i.e., social performance-based incentives combined with A&F), and hence we cannot separate between them. Accordingly, another exciting avenue for future

[25] For example, the literature on nonprofit governance typically focuses on issues that are germane to large nonprofits in economically developed countries (e.g., Bertrand et al. 2020, Fisman and Hubbard 2005, Krause et al. 2019). Similarly, the operations research literature in healthcare—which studies how to optimize hospitals' operating processes—has focused on large-scale hospitals in the U.S. and other high-income countries (Berry Jaeker and Tucker 2016, Roth, Tucker, and Venkataraman 2019, Song et al. 2018).

research would be to "un-bundle" the governance bundle and examine how individual governance practices contribute to non-profits' social impact. Third, and relatedly, future research could explore the effectiveness of other types of governance and management practices available to non-profit organizations and track their impact over time.

Lastly—and perhaps most importantly—our findings have important implications for practice. Every year, large amounts of funding are invested in non-profit organizations pursuing social and environmental causes and aiming to achieve the seventeen SDGs of the United Nations (e.g., ending poverty, reducing hunger, promoting healthy lives and well-being, reducing inequalities, addressing climate change, protecting life on land and below water). The insights of this study help inform non-profit organizations, as well as their donors and impact investors, about the governance mechanisms that are available and effective in achieving the non-profits' objectives and maximizing the social impact of the funds invested.

References

Adam T, Lim SS, Mehta S, Bhutta ZA, Fogstad H, Mathai M, Zupan J, Darmstadt GL (2005) Cost effectiveness analysis of strategies for maternal and neonatal health in developing countries. *British Medical J.* 331(7525):1107.

Aguilera R, Florackis C, Kim H (2016) Advancing the corporate governance research agenda. *Corporate Governance: An International Rev.* 24(3):172–180.

Amabile TM, Goldfarb P, Brackfield SC (1990) Social influences on creativity: evaluation, coaction, and surveillance. *Creativity Research J.* 3(1):6–21.

Anderson ML (2008) Multiple inference and gender differences in the effects of early intervention: a reevaluation of the abecedarian, Perry preschool, and early training projects. *J. American Stat. Association* 103(484):1481–1495.

Ashraf N, Bandiera O, Jack BK (2014) No margin, no mission? A field experiment on incentives for public service delivery. *J. Public Econom.* 120(6):1–17.

Ashraf N, Bandiera O, Lee S (2014) Awards unbundled: evidence from a natural field experiment. *J. Econom. Behav. Organ.* 100(4):44–63.

Ballesteros L, Gatignon A (2019) The relative value of firm and nonprofit experience: tackling large-scale social issues across institutional contexts. *Strategic Management J.* 40(4):631–657.

Bandiera O, Larcinese V, Rasul I (2015) Blissful ignorance? A natural experiment on the effect of feedback on students' performance. *Labour Econom.* 34(C):13–25.

Basinga P, Gertler PJ, Binagwaho A, Soucat AL, Sturdy J, Vermeersch CM (2011) Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. *Lancet* 377(9775):1421–1428.

Battilana J, Dorado S (2010) Building sustainable hybrid organizations: the case of commercial microfinance organizations. *Acad. Management J.* 53(6):1419–1440.

Berrone P, Gelabert L, Massa-Saluzzo F, Rousseau HE (2016) Understanding community dynamics in the study of grand challenges: how nonprofits, institutional actors, and the community fabric interact to influence income inequality. *Acad. Management J.* 59(6):1940–1964.

Berry Jaeker JA, Tucker AL (2016) Past the point of speeding up: the negative effects of workload saturation on efficiency and patient severity. *Management Sci.* 63(4):1042–1062.

Bertrand M, Bombardini M, Fisman R, Trebbi F (2020) Tax-exempt lobbying: corporate philanthropy as a tool for political influence. *American Econom. Rev.* 110(7):2065–2102.

Blader S, Gartenberg C, Prat A (2020) The contingent effect of management practices. *Rev. Economic Stud.* 87(2):721–749.

Bloom N, Brynjolfsson E, Foster L, Jarmin R, Patnaik M, Saporta-Eksten I, Van Reenen J (2019) What drives differences in management practices? *American Econom. Rev.* 109(5):1648–1683.

Bloom N, Eifert B, Mahajan A, McKenzie D, Roberts J (2013) Does management matter? Evidence from India. *Quart. J. Econom.* 128(1):1–51.

Bloom N, Van Reenen J (2007) Measuring and explaining management practices across firms and countries. *Quart. J. Econom.* 122(4):1351–1408.

Bode C, Singh J (2018) Taking a hit to save the world? Employee participation in a corporate social initiative. *Strategic Management J.* 39(4):1003–1030.

Bowles S (2016) *The Moral Economy: Why Good Incentives are no Substitute for Good Citizens* (Yale University Press, New Haven, CT).

Bruhn M, Karlan D, Schoar A (2018) The impact of consulting services on small and medium enterprises: evidence from a randomized trial in Mexico. *J. Political Economy* 126(2):635–687.

Bryan KA, Tilsick A, Zhu B (2017) Which entrepreneurs are coachable and why? *American Econom. Rev.* 107(5):312–316.

Burbano VC (2016) Social responsibility messages and worker wage requirements: field experimental evidence from online labor marketplaces. *Organ. Sci.* 27(4):1010–1028.

Cabral S, Mahoney JT, McGahan AM, Potoski M (2019) Value creation and value appropriation in public and nonprofit organizations. *Strategic Management J.* 40(4):465–475.

Cai J, Szeidl A (2018) Interfirm relationships and business performance. *Quart. J. Econom.* 133(3):1229–1282.

Carnahan S, Kryscynski D, Olson D (2017) When does corporate social responsibility reduce employee turnover? Evidence from attorneys before and after 9/11. *Acad. Management J.* 60(5):1932–1962.

Cassar L (2019) Job mission as a substitute for monetary incentives: benefits and limits. *Management Sci.* 65(2):896–912.

Cassar L, Meier S (2018) Nonmonetary incentives and the implications of work as a source of meaning. *J. Economic Perspectives* 32(3):215–238.

Cassar L, Meier S (2021) Intentions for doing good matter for doing well: the negative effects of prosocial incentives. *The Economic J.* 131(637):1988–2017.

Chatain O, Plaksenkova E (2019) NGOs and the creation of value in supply chains. *Strategic Management J.* 40(4):604–630.

Chatterji A, Delecourt S, Hasan S, Koning R (2019) When does advice impact startup performance? *Strategic Management J.* 40(3):331–356.

Cobb JA, Wry T, Zhao EY (2016) Funding financial inclusion: institutional logics and the contextual contingency of funding for microfinance organizations. *Acad. Management J.* 59(6):2103–2131.

De Mel S, McKenzie D, Woodruff C (2014) Business training and female enterprise start-up, growth, and dynamics: experimental evidence from Sri Lanka. *J. Development Econom.* 106(C):199–210.

Dimitriadis S, Koning R (2022) Social skills improve business performance: evidence from a randomized control trial with entrepreneurs in Togo. *Management Sci.*, forthcoming.

Dur R, Non A, Roelfsema H (2010) Reciprocity and incentive pay in the workplace. *J. Economic Psychology* 31(4):676–686.

Durand R, Huysentruyt M (2022) Communication frames and beneficiary engagement in corporate social initiatives: evidence from a randomized controlled trial in France. *Strategic Management J.*, forthcoming.

Eccles RG, Ioannou I, Serafeim G (2014) The impact of corporate sustainability on organizational processes and performance. *Management Sci.* 60(11):2835–2857.

Fisman R, Hubbard RG (2005) Precautionary savings and the governance of nonprofit organizations. *J. Public Econom.* 89(11–12):2231–2243.

Flammer C (2015) Does corporate social responsibility lead to superior financial performance? A regression discontinuity approach. *Management Sci.* 61(11):2549–2568.

Flammer C, Hong B, Minor DB (2019) Corporate governance and the rise of integrating corporate social responsibility criteria in executive compensation: effectiveness and implications for firm outcomes. *Strategic Management J.* 40(7):1097–1122.

Flammer C, Luo J (2017) Corporate social responsibility as an employee governance tool: evidence from a quasi-experiment. *Strategic Management J.* 38(2):163–183.

Gallus J, Frey BS (2016) Awards: a strategic management perspective. *Strategic Management J.* 37(8):1699–1714.

George G, Howard-Grenville J, Joshi A, Tihanyi L (2016a) Understanding and tackling societal grand challenges through management research. *Acad. Management J.* 59(6):1880–1895.

George G, Corbishley C, Khayesi JN, Haas MR, Tihanyi L (2016b). Bringing Africa in: promising directions for management research. *Acad. Management J.* 59(2):377–393.

Giorcelli M (2019) The long-term effects of management and technology transfers. *American Econom. Rev.* 109(1):121–152.

Governance Institute (2018) *The Distinction between Management and Governance*, 2nd edition (Governance Institute, Lincoln, NE).

Gubler T, Larkin I, Pierce L (2016) Motivational spillovers from awards: crowding out in a multitasking environment. *Organ. Sci.* 27(2):286–303.

Gubler T, Larkin I, Pierce L (2017) Doing well by making well: the impact of corporate wellness programs on employee productivity. *Management Sci.* 64(11):4967–4987.

Hansmann HB (1980) The role of nonprofit enterprise. *Yale Law J.* 89(5):835–901.

Hanson D, Hitt MA, Ireland RD, Hoskisson RE (2017) *Strategic Management: Competitiveness and Globalisation*, 6th edition (Cengage, Boston, MA).

Hollowell J, Oakley L, Kurinczuk JJ, Brocklehurst P, Gray R (2011) The effectiveness of antenatal care programmes to reduce infant mortality and preterm birth in socially disadvantaged and vulnerable women in high-income countries: a systematic review. *BMC Pregnancy and Childbirth* 11(1):13.

Holmstrom B, Milgrom P (1991) Multitask principal-agent analyses: incentive contracts, asset ownership, and job design. *J. Law Econom. Organ.* 7:24–52.

Huillery E, Seban J (2021) Financial incentives, efforts, and performances in the health sector: experimental evidence from the Democratic Republic of Congo. *Econom. Development and Cultural Change* 69(3):1115–1164.

Jay J (2013) Navigating paradox as a mechanism of change and innovation in hybrid organizations. *Acad. Management J.* 56(1):137–159.

Kellogg KC (2009) Operating room: relational spaces and micro-institutional change in surgery. *American J. of Sociology* 115(3):657–711.

King AA, Lenox MJ (2001) Does it really pay to be green? An empirical study of firm environmental and financial performance. *J. Industrial Ecology* 5(1):105–116.

Krause R, Zhiyan WU, Bruton GD, Carter SM (2019) The coercive isomorphism ripple effect: an investigation of nonprofit interlocks on corporate boards. *Acad. Management J.* 62(1):283–308.

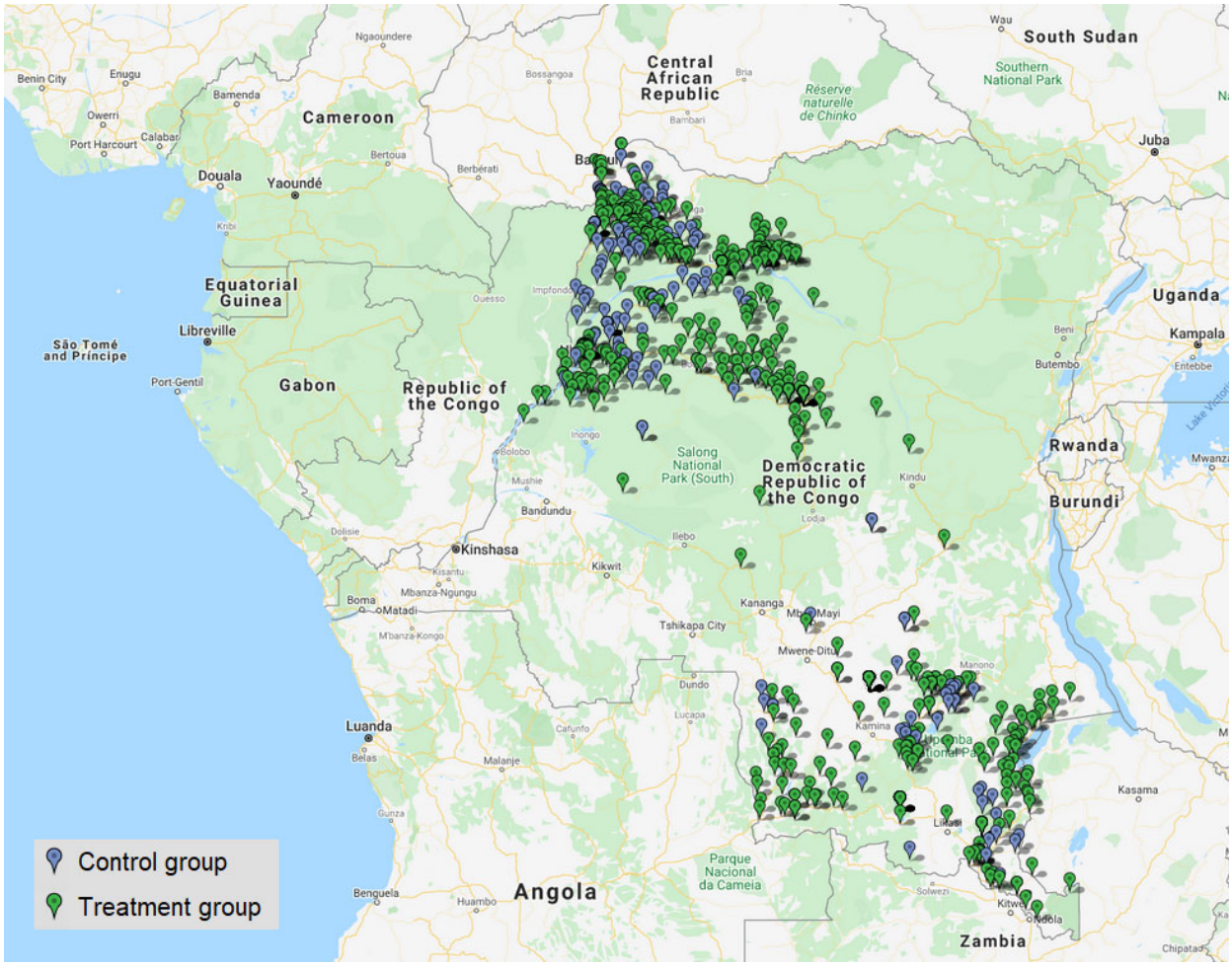
Linden LL, Shastry GK (2012) Grain inflation: identifying agent discretion in response to a conditional school nutrition program. *J. Development Econom.* 99(1):128–138.

Mair J, Marti I, Ventresca MJ (2012) Building inclusive markets in rural Bangladesh: how intermediaries work institutional voids. *Acad. Management J.* 55(4):819–850.
Mayaka MS, Muvudi Lushimba M, Bertone MP, de Borman N (2011) Le financement basé sur la performance en République Démocratique du Congo: comparaison de deux expériences pilotes. PBF CoP Working Paper WP6 (Performance-based Financing Community of Practice, Antwerp, Belgium).
McKenzie D (2021) Small business training to improve management practices in developing countries: reassessing the evidence for "training doesn't work." *Oxford Rev. Econom. Policy* 37(2):276–301.
McKenzie D, Woodruff C (2014) What are we learning from business training evaluations around the developing world? *World Bank Research Observer* 29(1):48–82.
McKenzie D, Woodruff C (2017) Business practices in small firms in developing countries. *Management Sci.* 63(9):2967–2981.
Pache AC, Santos F (2013) Inside the hybrid organization: selective coupling as a response to competing institutional logics. *Acad. Management J.* 56(4):972–1001.
PDSS (2016) *Manuel Opérationnel du Financement Basé sur la Performance* (Ministry of Health, Democratic Republic of Congo).
Quélin BV, Kivleniece I, Lazzarini S (2017) Public-private collaboration, hybridity and social value: towards new theoretical perspectives. *J. Management Stud*. 54(6):763–792.
Roth A, Tucker AL, Venkataraman S (2019) Being on the productivity frontier: identifying "triple aim performance" hospitals. *Production and Operations Management* 28(9):2165–2183.
Rousseau HE, Berrone P, Gelabert L (2019) Localizing sustainable development goals: nonprofit density and city sustainability. *Acad. Management Discoveries* 5(4):487–513.
Smither JW, London M, Reilly RR (2005) Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Personnel Psych.* 58(1):33–66.
Song H, Tucker AL, Murrel KL, Vinson DR (2018) Closing the productivity gap: improving worker productivity through public relative performance feedback and validation of best practices. *Management Sci*. 64(6):2628–2649.
UNICEF (2016) *Maternal and Newborn Health Disparities: Democratic Republic of the Congo* (UNICEF, New York, NY).
United Nations (2018a) *Satellite Account on Non-profit and Related Institutions and Volunteer Work* (The United Nations, New York, NY).
United Nations (2018b) *Levels & Trends in Child Mortality* (The United Nations Inter-Agency Group for Child Mortality Estimation, New York, NY).
Vakili K, McGahan AM (2016) Health care's grand challenge: stimulating basic science on diseases that primarily afflict the poor. *Acad. Management J.* 59(6):1917–1939.
Viscusi WK, Masterman CJ (2017) Income elasticities and global values of a statistical life. *J. Benefit-Cost Anal.* 8(2):226–250.
WHO (2015) *Democratic Republic of the Congo: Improving Aid Coordination in the Health Sector* (World Health Organization, Geneva).
WHO (2018) *Children: Reducing Mortality* (World Health Organization, Geneva).
WHO (2019a) *Stillbirths* (World Health Organization, Geneva).
WHO (2019b) *Neonatal Mortality* (World Health Organization, Geneva).
World Bank (2018) *Health System Strengthening for Better Maternal and Child Health Results Project* (World Bank, Washington, DC).
Wrzesniewski A, Schwartz B, Cong X, Kane M, Omar A, Kolditz T (2014) Multiple types of motives don't multiply the motivation of West Point cadets. *Proceedings of the National Acad. of Sciences* 111(30):10990–10995.

Figure 1. Examples of health centers

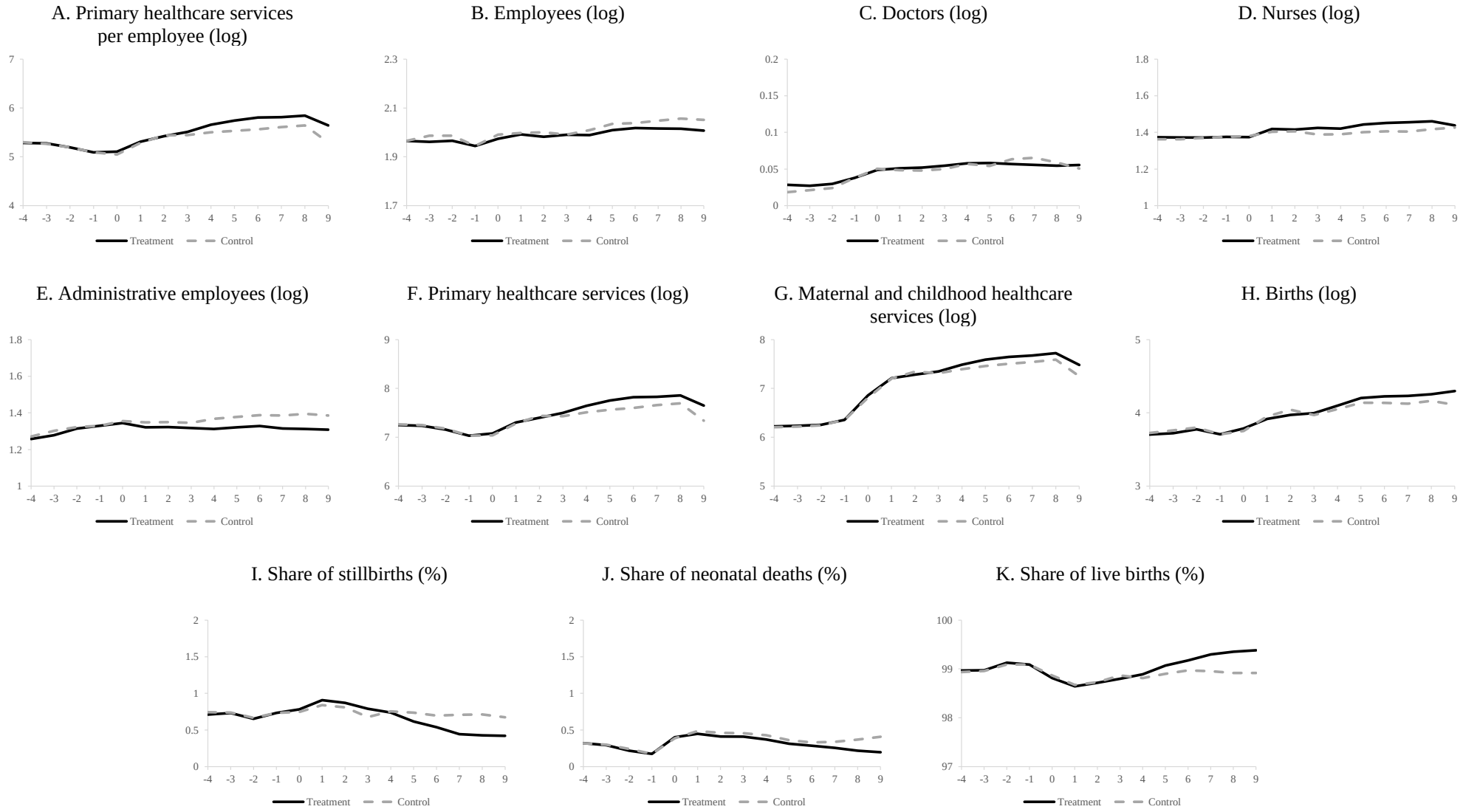
Source: Bluesquare.

Figure 2. Location of treated and control health centers



Notes. This figure plots the location of the health centers in the control group (blue markers) and treatment group (green markers), for the health centers in our sample that have non-missing geo-codes (73% of the sample). The map was generated using the online mapping software Maptive.

Figure 3. Dynamics



Notes. This figure plots the evolution of each outcome variable in the treatment and control groups. In each panel, the *y*-axis refers to the mean of the respective outcome variable across all health centers in the treatment (black solid line) and control (gray dashed line) groups, respectively. The *x*-axis indicates quarters in event time, where $t = 0$ is the initial quarter of the PDSS intervention (Q1 2017). In each plot, the mean of the control group is normalized to match the mean of the treatment group at $t = -1$.

Table 1. Summary of the PDSS intervention

	Treated health centers	Control health centers
Performance-based incentives	• Performance is assessed along 22 healthcare performance indicators (Table A1) and a quality score of up to 400 points based on 15 criteria (Table A2). • The performance-based subsidy is then computed according to equation (3).	–
Auditing & feedback	Auditing • Auditors review the number of services provided at the health center and service quality based on the assessment grid used to determine the performance-based payments. • Auditors are assisted by local associations whose mandate is to cross-check the existence of the patients indicated in the health centers' registries and collect information about patients' satisfaction ("community verification system"). Feedback • Auditors provide feedback and recommendations (guided by the discrepancies between current and target performance); they support the write-up of the quarterly management plan that outlines how the improvements can be achieved; and they provide training in financial and stock management.	–
Transfers	• Initial subsidy ("unité d'investissement") at the beginning of the program. • Quarterly performance-based subsidy computed according to equation (3).	• Initial subsidy ("unité d'investissement") at the beginning of the program. • Quarterly subsidy that matches the average subsidy paid to the treated health centers in the same quarter and the same health district. This subsidy does *not* depend on the control health centers' own performance.

Source: PDSS (2016).

Table 2. Summary statistics

	Obs.	Mean	Median	Std. Dev.
Panel A. Health centers statistics				
Primary healthcare services per employee	999	235.76	184.88	214.01
Employees	999	7.13	6	6.30
Doctors	999	0.06	0	0.31
Nurses	999	3.55	3	3.70
Administrative	999	3.53	3	3.37
Primary healthcare services	999	1,611	1,344	1,195
Maternal and childhood healthcare services	999	957	705	940
Births	999	53.80	45	39.46
Stillbirths (in %)	999	0.70	0	1.51
Neonatal deaths (in %)	999	0.17	0	0.68
Live births (in %)	999	99.12	100	1.76
Panel B. Population statistics				
Population in center's health area	999	11,135	9,508	7,734
Population in center's health district	999	204,409	181,565	80,683

Notes. All variables are recorded in the quarter preceding the PDSS intervention (that is, Q4 2016).

Table 3. Randomization tests

		Obs.	Mean	Median	Std. Dev.	*p*-value (diff. means)	*p*-value (KS test)
Panel A. Health centers statistics							
Primary healthcare services per employee	Treated	674	238.53	176.62	231.07	0.780	0.156
	Control	325	230.02	200.29	173.55		
Employees	Treated	674	7.43	6	7.21	0.196	0.140
	Control	325	6.50	6	3.67		
Doctors	Treated	674	0.06	0	0.33	0.457	0.885
	Control	325	0.04	0	0.25		
Nurses	Treated	674	3.79	3	4.26	0.135	0.675
	Control	325	3.05	3	2.03		
Administrative	Treated	674	3.58	3	3.66	0.710	0.453
	Control	325	3.42	3	2.68		
Primary healthcare services	Treated	674	1,660	1,361	1,273	0.351	0.213
	Control	325	1,508	1,338	1,009		
Maternal and childhood healthcare services	Treated	674	988	710	1,010	0.494	0.729
	Control	325	892	684	770		
Births	Treated	674	55.71	45	41.42	0.348	0.342
	Control	325	49.84	45	34.78		
Stillbirths (in %)	Treated	674	0.73	0	1.53	0.500	0.871
	Control	325	0.64	0	1.45		
Neonatal deaths (in %)	Treated	674	0.18	0	0.68	0.981	0.980
	Control	325	0.17	0	0.68		
Live births (in %)	Treated	674	99.09	100	1.772	0.584	0.863
	Control	325	99.18	100	1.749		
Panel B. Population statistics							
Population in center's health area	Treated	674	11,090	9,491	6,377	0.896	0.776
	Control	325	11,227	9,847	9,988		
Population in center's health district	Treated	674	201,829	181,565	73,925	0.716	0.458
	Control	325	209,760	177,275	93,047		

Notes. All variables are recorded in the quarter preceding the PDSS intervention (that is, Q4 2016). The penultimate column reports the *p*-value of the difference-in-means test comparing treated and control health centers. The last column reports the *p*-value of the Kolmogorov-Smirnov (KS) test of identical distributions comparing treated and control health centers.

Table 4. The impact of the governance treatment on health center outcomes

	Health center operating efficiency	Health center employees				Volume of healthcare services			Quality of healthcare services		
	Primary healthcare services per employee (log)	Employees (log)	Doctors (log)	Nurses (log)	Administrative employees (log)	Primary healthcare services (log)	Maternal and childhood healthcare services (log)	Births (log)	Share of stillbirths (%)	Share of neonatal deaths (%)	Share of live births (%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Treatment × post (t = 0 to 4)	0.068	-0.003	-0.004	0.012	-0.021	0.065	0.017	0.023	0.065	-0.031	-0.034
	(0.042)	(0.016)	(0.008)	(0.028)	(0.023)	(0.045)	(0.062)	(0.040)	(0.069)	(0.052)	(0.095)
Treatment × post (t = 5 to 9)	0.255***	-0.022	-0.008	0.032	-0.058**	0.233***	0.149*	0.138***	-0.209***	-0.107**	0.316***
	(0.059)	(0.021)	(0.009)	(0.036)	(0.029)	(0.068)	(0.077)	(0.048)	(0.072)	(0.053)	(0.099)
Health center fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.561	0.785	0.739	0.643	0.769	0.523	0.565	0.564	0.290	0.323	0.355
Observations	13,986	13,986	13,986	13,986	13,986	13,986	13,986	13,986	13,986	13,986	13,986

Notes. The sample period is from Q1 2016 (t = – 4) until Q2 2019 (t = 9), where Q1 2017 (t = 0) is the quarter of the PDSS intervention. Standard errors (reported in parentheses) are block-bootstrapped at the health center level using 1,000 bootstrap replications. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.

Table 5. Dynamics

	Health center operating efficiency	Health center employees				Volume of healthcare services			Quality of healthcare services		
	Primary healthcare services per employee (log)	Employees (log)	Doctors (log)	Nurses (log)	Administrative employees (log)	Primary healthcare services (log)	Maternal and childhood healthcare services (log)	Births (log)	Share of stillbirths (%)	Share of neonatal deaths (%)	Share of live births (%)
Treatment × pre (t = –4)	-0.023	0.008	0.011	0.011	-0.009	-0.015	0.016	-0.013	-0.043	0.000	0.042
	(0.035)	(0.021)	(0.009)	(0.019)	(0.032)	(0.029)	(0.056)	(0.056)	(0.103)	(0.062)	(0.130)
Treatment × pre (t = –3)	0.012	-0.024	0.005	0.012	-0.029	-0.012	0.018	-0.043	-0.005	0.002	0.003
	(0.035)	(0.019)	(0.008)	(0.018)	(0.029)	(0.032)	(0.054)	(0.057)	(0.103)	(0.062)	(0.128)
Treatment × pre (t = –2)	-0.001	-0.028	0.008	0.001	-0.011	-0.028	0.017	-0.030	-0.016	-0.031	0.048
	(0.034)	(0.020)	(0.008)	(0.017)	(0.029)	(0.031)	(0.057)	(0.051)	(0.115)	(0.067)	(0.139)
Treatment × post (t = 0)	0.079	-0.022	-0.002	-0.008	-0.015	0.057	0.066	0.045	0.045	0.022	-0.067
	(0.059)	(0.020)	(0.009)	(0.021)	(0.025)	(0.053)	(0.070)	(0.056)	(0.095)	(0.070)	(0.126)
Treatment × post (t = 1)	0.006	-0.000	0.004	0.024	-0.032	0.005	-0.010	-0.061	0.075	-0.057	-0.018
	(0.060)	(0.022)	(0.007)	(0.025)	(0.024)	(0.069)	(0.083)	(0.050)	(0.126)	(0.071)	(0.153)
Treatment × post (t = 2)	-0.025	-0.025	0.004	0.004	-0.026	-0.050	-0.081	-0.074	0.056	-0.047	-0.009
	(0.061)	(0.024)	(0.007)	(0.028)	(0.023)	(0.069)	(0.081)	(0.055)	(0.118)	(0.073)	(0.150)
Treatment × post (t = 3)	0.095	0.007	0.005	0.047	-0.029	0.102	0.072	0.057	0.132	-0.047	-0.085
	(0.069)	(0.027)	(0.008)	(0.034)	(0.024)	(0.081)	(0.095)	(0.064)	(0.122)	(0.073)	(0.155)
Treatment × post (t = 4)	0.173**	-0.029	-0.001	0.025	-0.064**	0.144*	0.100	0.043	-0.066	-0.061	0.127
	(0.069)	(0.028)	(0.008)	(0.034)	(0.025)	(0.080)	(0.090)	(0.062)	(0.130)	(0.073)	(0.156)
Treatment × post (t = 5)	0.227***	-0.024	0.006	0.048	-0.055**	0.203***	0.141	0.069	-0.140	-0.042	0.182
	(0.066)	(0.028)	(0.008)	(0.036)	(0.026)	(0.079)	(0.087)	(0.060)	(0.120)	(0.068)	(0.145)
Treatment × post (t = 6)	0.241***	-0.018	-0.011	0.044	-0.061**	0.223***	0.139	0.097	-0.163	-0.044	0.207
	(0.064)	(0.026)	(0.009)	(0.032)	(0.026)	(0.076)	(0.087)	(0.060)	(0.116)	(0.068)	(0.141)
Treatment × post (t = 7)	0.188***	-0.036	-0.009	0.054	-0.074***	0.152**	0.131	0.109	-0.299**	-0.091	0.390***
	(0.064)	(0.027)	(0.010)	(0.035)	(0.028)	(0.076)	(0.086)	(0.066)	(0.119)	(0.069)	(0.148)
Treatment × post (t = 8)	0.206***	-0.044	-0.003	0.040	-0.086***	0.163**	0.133	0.083	-0.279**	-0.169**	0.448***
	(0.068)	(0.027)	(0.009)	(0.036)	(0.028)	(0.082)	(0.089)	(0.064)	(0.124)	(0.066)	(0.152)
Treatment × post (t = 9)	0.399***	-0.044	0.007	0.003	-0.075***	0.355***	0.267**	0.225***	-0.243**	-0.224***	0.467***
	(0.095)	(0.031)	(0.008)	(0.038)	(0.029)	(0.111)	(0.112)	(0.070)	(0.108)	(0.078)	(0.146)
Health center fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.562	0.785	0.739	0.643	0.769	0.523	0.565	0.564	0.290	0.324	0.356
Observations	13,986	13,986	13,986	13,986	13,986	13,986	13,986	13,986	13,986	13,986	13,986

Notes. The sample period is from Q1 2016 (t = – 4) until Q2 2019 (t = 9), where Q1 2017 (t = 0) is the quarter of the PDSS intervention. Standard errors (reported in parentheses) are block-bootstrapped at the health center level using 1,000 bootstrap replications. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.

Table 6. Quality improvements at the treated health centers

	Quality score $t = 0$ (Q1 2017)	Quality score $t = 9$ (Q2 2019)	Difference in means	p-value
Overall organization	15.86	66.52	50.66	0.000
Management plan	2.71	75.31	72.60	0.000
Financial management	7.87	62.88	55.01	0.000
Indigent committee	0.77	51.41	50.64	0.000
Hygiene and sterilization	13.07	45.57	32.50	0.000
External consultations	21.32	59.03	37.71	0.000
Family planning	8.92	47.17	38.25	0.000
Laboratory	19.75	40.58	20.83	0.000
Screening services	7.35	38.64	31.29	0.000
Drugs and consumables	17.84	65.29	47.45	0.000
Maternity care	21.95	63.17	41.22	0.000
Vaccination	31.63	61.63	30.00	0.000
Pre-natal care	32.11	71.50	39.39	0.000
HIV and tuberculosis care	6.41	22.32	15.91	0.000

Notes. This table reports the mean of the treated health centers' quality scores at $t = 0$ (Q1 2017) and $t = 9$ (Q2 2019), along with the difference in means and the corresponding p-value. The quality scores are described in Table A3. The sample consists of 646 treated health centers with non-missing quality scores.

Table 7. Quality of healthcare services at other healthcare facilities in the same health district as the treated health centers

	Hospitals in same health district as treated health centers			Control health centers in same health district as treated health centers			Outside health centers in same health district as treated health centers		
	Share of stillbirths (%)	Share of neonatal deaths (%)	Share of live births (%)	Share of stillbirths (%)	Share of neonatal deaths (%)	Share of live births (%)	Share of stillbirths (%)	Share of neonatal deaths (%)	Share of live births (%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Post (t = 0 to 4)	0.020	0.007	-0.027	-0.097	0.123	-0.026	0.024	0.006	-0.030
	(0.184)	(0.094)	(0.219)	(0.177)	(0.124)	(0.246)	(0.024)	(0.020)	(0.034)
Post (t = 5 to 9)	0.032	0.036	-0.067	-0.064	0.110	-0.046	0.033	-0.003	-0.030
	(0.281)	(0.145)	(0.320)	(0.198)	(0.172)	(0.293)	(0.041)	(0.033)	(0.057)
Health facility fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter fixed effects	No	No	No	No	No	No	No	No	No
R-squared	0.585	0.425	0.616	0.300	0.359	0.394	0.454	0.319	0.462
Observations	1,694	1,694	1,694	1,008	1,008	1,008	16,688	16,688	16,688

Notes. This table examines the changes in the share of stillbirths, neonatal deaths, and live births across healthcare facilities (hospitals in columns (1)-(3), control health centers in columns (4)-(6), and health centers outside the PDSS program in columns (7)-(9)) that are located within the same health district as the treated health centers. Standard errors (reported in parentheses) are block-bootstrapped at the health center level using 1,000 bootstrap replications. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.

Table 8. Funding vs. governance

	Health center operating efficiency	Health center employees				Volume of healthcare services			Quality of healthcare services		
	Primary healthcare services per employee (log)	Employees (log)	Doctors (log)	Nurses (log)	Administrative employees (log)	Primary healthcare services (log)	Maternal and childhood healthcare services (log)	Births (log)	Share of stillbirths (%)	Share of neonatal deaths (%)	Share of live births (%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Control vs. outside (t = 0 to 4)	-0.028	0.049	0.009	0.013	0.032	0.020	0.104	0.051	-0.079	0.015	0.064
	(0.069)	(0.032)	(0.010)	(0.032)	(0.047)	(0.066)	(0.126)	(0.061)	(0.131)	(0.062)	(0.157)
Treatment vs. outside (t = 0 to 4)	0.034	0.044	0.001	0.010	0.008	0.078	0.139	0.069	-0.047	-0.025	0.073
	(0.056)	(0.030)	(0.008)	(0.027)	(0.042)	(0.055)	(0.103)	(0.065)	(0.086)	(0.049)	(0.109)
Control vs. outside (t = 5 to 9)	0.209	0.178***	0.010	0.109**	0.115**	0.387**	0.470**	0.219**	-0.068	-0.080	0.148
	(0.173)	(0.049)	(0.010)	(0.049)	(0.057)	(0.194)	(0.220)	(0.109)	(0.115)	(0.064)	(0.143)
Treatment vs. outside (t = 5 to 9)	0.451***	0.150***	0.002	0.122***	0.046	0.601***	0.567***	0.340***	-0.223***	-0.198***	0.421***
	(0.126)	(0.046)	(0.008)	(0.041)	(0.052)	(0.145)	(0.165)	(0.079)	(0.083)	(0.056)	(0.117)
Health center fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Matched pairs by quarter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.527	0.649	0.704	0.644	0.607	0.552	0.553	0.582	0.258	0.241	0.299
Observations	27,972	27,972	27,972	27,972	27,972	27,972	27,972	27,972	27,972	27,972	27,972

Notes. This table presents variants of the regressions in Table 4, except that the treatment and control group are compared to health centers outside of the PDSS program ("outside" group) Health centers from the outside group are matched to health centers in the treatment and control groups, respectively, using the nearest-neighbor matching described in Section 6. Standard errors (reported in parentheses) are block-bootstrapped at the health center level using 1,000 bootstrap replications. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.