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

ANICET A. FANGWA	CAROLINE FLAMMER	MARIEKE HUYSENTRUYT	BERTRAND V. QUÉLIN
HEC Paris	Columbia University and NBER	HEC Paris	HEC Paris
anicet.fangwa@hec.edu	caroline.flammer@columbia.edu	huysentruyt@hec.fr	quelin@hec.fr

February 2023

ABSTRACT

Substantial funding is provided to the healthcare systems of low-income countries. However, an important challenge is to ensure that this funding be used efficiently. This challenge is complicated by the fact that a large share of healthcare services in low-income countries is provided by non-profit health centers that often lack i) effective governance structures and ii) organizational know-how and adequate training. In this paper, we argue that the bundling of performance-based incentives with auditing and feedback (A&F) is a potential way to overcome these obstacles. First, the combination of feedback and performance-based incentives—that is, feedback joint with incentives to act on this feedback and achieve specific health outcomes—helps address the knowledge gap that may otherwise undermine performance-based incentives. Second, coupling feedback with auditing helps ensure that the information underlying the feedback is reliable—a prerequisite for effective feedback. To examine the effectiveness of this bundle, we use data from a randomized governance program conducted in the Democratic Republic of Congo. Within the program, a set of health centers were randomly assigned to a "governance treatment" that consisted of performance-based incentives combined with A&F, while others were not. Consistent with our prediction, we find that the governance treatment led to i) higher operating efficiency and ii) improvements in health outcomes. Furthermore, we find that funding is not a substitute for the governance treatment—health centers that only receive funding increase their scale, but do not show improvements in operating efficiency nor health outcomes.

Keywords: non-profit governance; non-profit organizations; social impact; healthcare; low-income countries; randomized experiment.

* We are grateful to the Editor (Lamar Pierce), an anonymous Associate Editor, three anonymous reviewers, as well as Ruth Aguilera, Dominique Baabo Kubuya, Marco Becht, Vanessa Burbano, Olga Hawn, Natasha Hochberg, David Levine (AEA discussant) Steve Lydenberg, Serge Mayaka, Mary-Hunter McDonnell, Agnes Muzama, Ivan Png, Gui Deng Say, David Stolow, Haitao Yu, as well as seminar participants at Boston University, Cornell University, Erasmus 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 AEA Meetings (New Orleans), 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).

1. Introduction

Over the past decades, large amounts of funding have been dedicated to the improvement of national healthcare systems in low-income countries such as the Democratic Republic of Congo (DRC). However, the recent literature has questioned the effectiveness of these funds—which are typically provided in the form of lump-sum transfers or performance-based financing—in improving the quality of healthcare services (e.g., Eichler and Levine 2009, Huillery and Seban 2021, Linden and Shastry 2012, Miller and Babiarz 2013, Paul et al. 2018). In this regard, an important challenge is the fact that a large share of healthcare services in low-income countries is provided by non-profit health centers that often lack effective governance structures, organizational know-how, and adequately trained staff.[1]

In this paper, we examine potential governance mechanisms through which non-profit health centers in low-income countries can mitigate these obstacles and ultimately improve the quality of their services. Specifically, we propose that the bundling of i) performance-based funding and ii) auditing and feedback (A&F) is a way to address these obstacles. The rationale is as follows. First, the sole use of performance-based funding is unlikely to be effective if the non-profit's employees lack the necessary knowledge on how to improve the organization's operations, and may even backfire if the non-profit's employees feel overwhelmed by the performance targets. As we argue in this paper, the combination of performance-based incentives and feedback—that is, feedback joint with incentives to act on this feedback and achieve specific health outcomes—helps address the knowledge gap that may otherwise undermine performance-based incentives. Moreover, the provision of effective feedback requires reliable information—another important challenge in low-income countries (e.g., Kerber et al. 2015, Kumar et al. 2022)—which can be addressed by further combining feedback with auditing.

[1] These obstacles are commonly seen as a critical threat to the development of viable healthcare systems in low-income countries (e.g., WHO 2021, World Bank 2018). For example, in their study of the governance practices of healthcare facilities in low- and middle-income countries, Saif-Ur-Rahman et al. (2019) document a large "governance gap," highlighting that "[g]overnance is one of the most important aspects for strong primary healthcare (PHC) service delivery. To achieve the targets for the Sustainable Development Goals, good governance may play a prime role in low-income and middle-income countries" (p. 1).

By considering the bundling of performance-based funding with A&F, we contribute to the literature that studies the design of effective governance practices. Researchers in economics, strategy, and organization have long studied how different governance practices can improve firm performance (for reviews of this literature, see Aguilera et al. 2016, Bebchuk and Weisbach 2010, Becht, Bolton, and Roell 2003). However, this literature offers only limited guidance in our context, as it focuses primarily on the governance practices of for-profit organizations in high-income countries, which need not be effective for non-profits in low-income countries. For example, the different objective function of non-profits, along with the lack of organizational know-how, adequately trained employees, and verifiable information may undermine the use of traditional governance tools. And indeed, the existing literature finds that performance-based funding is ineffective in improving—and sometimes even *reduces*—the quality of healthcare services in low-income countries (e.g., Basinga et al. 2011, Huillery and Seban 2021, Linden and Shastry 2012, Mayaka et al. 2011).[2] This is puzzling and underscores the challenge of how to think about and design effective governance mechanisms for non-profits in low-income countries. As we argue, a potential remedy is to combine performance-based funding with other, complementary provisions—namely auditing and feedback—that take into account the nature of non-profits and the specific challenges that they face in low-income countries. In this regard, our paper echoes Milgrom and Roberts' (1988, 1990, 1995) insight that business practices can be substantially enhanced when complementarities exist between these practices.

To empirically examine whether the bundling of performance-based funding with A&F can improve health outcomes in low-income countries, we use detailed micro data obtained from a randomized governance program that was conducted by the World Bank in the DRC's healthcare sector. This program, which aimed to increase the quantity and quality of primary healthcare services—and especially maternity and childbirth services—was administered in about 1,000 health centers starting at the beginning of 2017.

[2] A similar conclusion was reached by practitioners. For example, a recent report of the Independent Evaluation Group (IEG) of the World Bank highlights the disappointing results of performance-based funding in the healthcare sector of low-income countries (IEG 2014). See also the related discussion in Paul et al. (2018).

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 performance-based incentives combined with A&F. Hence, by design, this randomized governance program provides an ideal setup to study how this bundle of governance practices affects health centers' outcomes (e.g., their operating efficiency and service quality), holding everything else, including funding, constant.[3]

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 indicate that the intervention—that is, the bundling of performance-based funding with A&F—helps the treated health centers improve their operating efficiency and increase the quality of their services.

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 and achieve specific health outcomes—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.[4]

---

[3] 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. 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." That being 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 centers' daily operations are conducted. 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).

[4] 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

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 auditing. 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 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, in auxiliary analyses, we examine whether funding can serve as a substitute for the governance treatment. 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 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 the governance treatment—health centers that only receive funding increase their scale, but do not show improvements in

learn and adopt best practices from one's 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.

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.

Naturally, we caution that our findings are specific to the healthcare sector in a low-income country, and hence need not generalize to other settings. Nevertheless, studying this specific context is important, as improving children's health has been an important challenge in low-income countries. 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. In particular, 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 2018). Reducing children mortality is of foremost importance, and this study helps inform how specific governance practices can contribute to this objective.

Our study makes several contributions to the literature. First, it contributes to the governance literature by exploring how the bundling of performance-based funding and A&F can increase the effectiveness of non-profit organizations in low-income countries. Second, it adds to the literature on the "bundling" of business practices, highlighting how performance-based funding and A&F can complement each other in addressing the key obstacles faced by non-profit health centers in low-income countries. Third, it advances the literature that studies the effectiveness of performance-based incentives in low-income countries, and provides a potential explanation for their limited effectiveness, along with a potential remedy.[5]

[5] In Section 7, we elaborate on these contributions and discuss the related literature.

2. Theoretical background

The challenge of designing effective governance practices for non-profits in low-income countries

A large literature examines how different governance practices can improve the performance of for-profit organizations in high-income countries (e.g., Aguilera et al. 2016, Bebchuk and Weisbach 2010, Becht, Bolton, and Roell 2003). While related, the insights from this literature need not generalize to non-profits in low-income countries. Indeed, in order to be effective, appropriate governance practices need to take into account the unique nature of non-profits in low-income countries and the obstacles that they face. First, unlike for-profit organizations, non-profits aim to maximize their social impact (such as better health outcomes) as opposed to financial performance. Accordingly, traditional governance mechanisms—such as pay-for-financial performance or profit-sharing—may not be suitable, as they are not necessarily aligned with the non-profits' objectives (Glaeser 2002, Hansmann 1980). Second, non-profit employees are often purpose-driven and intrinsically motivated. As such, they are likely less responsive to extrinsic incentives (Leete 2000). In fact, the provision of extrinsic incentives could even crowd out their intrinsic motivation and ultimately hurt the non-profit's objective.[6] Third, non-profits in low-income countries operate in an environment that is resource constrained. In particular, their managers and employees may lack the necessary knowledge on how to improve the organizations' operations.[7]

As a way to mitigate these challenges, a potentially appealing incentive mechanism is the provision of performance-based funding, in which non-profit organizations receive additional funding conditional on meeting specific social goals (e.g., specific health outcomes). As such, performance-based funding differs from the more traditional pay-for-financial performance incentives in two ways: i) the additional funding is tied to social criteria in lieu of financial criteria, and ii) the direct beneficiary is the non-profit organization

[6] This argument echoes the findings of studies in the context of for-profit organizations that highlight a potential "dark side" of financial performance-based incentives, in that such incentives can potentially crowd out the employee's intrinsic motivation and ultimately reduce their engagement and job satisfaction (e.g., Bowles 2016, Cassar and Meier 2021, Gubler, Larkin, and Pierce 2016, Wrzesniewski et al. 2014).

[7] For example, in the context of the DRC's healthcare sector, the WHO has identified several sources of inefficiencies such as the lack of strategic and managerial planning, inadequate priorities in resource allocation, lack of managerial competencies, and insufficient medical training of health workers (WHO 2015).

itself as opposed to the managers and employees, respectively.

The provision of performance-based funding is appealing, as it is aligned with the non-profit's social objective, and may motivate the non-profit's managers and employees in two ways. First, by providing such incentives, non-profits can leverage their managers' and employees' intrinsic motivation to obtain additional funding for the non-profit's cause.[8] Second, performance-based funding 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.[9] On top of the motivational aspect, providing employees and managers with training, upgraded equipment, etc., may empower them to further increase their productivity.

While the use of performance-based funding is appealing from a theoretical perspective, the empirical literature finds that performance-based funding tends to be ineffective in low-income countries (e.g., Eichler and Levine 2009, Linden and Shastry 2012, Miller and Babiarz 2013). In fact, previous studies find that performance-based funding can even reduce the quality of healthcare services in low-income countries (e.g., Huillery and Seban 2021). This is puzzling and underscores the challenge of how to think about and design effective governance mechanisms for non-profits in low-income countries.

A potential explanation for the ineffectiveness of performance-based funding could be the lack of adequate training and organizational know-how in low-income countries. Arguably, the provision of performance-based funding may overwhelm and demotivate employees who lack the necessary knowledge to improve the non-profit's performance, which in turn could hurt (rather than improve) their performance

[8] 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).

[9] 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).

(Huillery and Seban 2021). Moreover, compared to financial performance metrics, non-financial performance metrics are difficult to measure and hence likely less suitable as a basis for performance-based incentives.[10] They might also be subject to manipulations absent proper auditing, which is often considered a pervasive issue in low-income countries (Linden and Shastry 2012). Relatedly, if multiple tasks contribute to the organization's objective, and if only selected tasks are part of the performance targets (such as those for which it is easier to obtain verifiable information), performance-based incentives may encourage substitution away from the non-incentivized activities, reorienting the employees' attention toward the incentivized activities at the expense of other activities that may have been essential to the organization's objective (Basinga et al. 2011, Holmstrom and Milgrom 1991).

The bundling of performance-based funding and A&F

As the above considerations illustrate, designing effective governance practices for non-profits in low-income countries is challenging due to their unique nature and the obstacles that they face. In what follows, we argue that the bundling of i) performance-based funding and ii) A&F is a potential way to overcome these challenges, as they complement each other in addressing the key obstacles faced by non-profits in low-income countries. These complementarities arise in three ways.

First, a key challenge of performance-based incentives is that they may overwhelm and demotivate employees who lack the necessary knowledge to improve the non-profit's operations, a challenge that is especially severe given the lack of organizational know-how and adequate training in low-income countries. The provision of feedback offers a direct remedy to this challenge—that is, the bundling of feedback with performance-based funding can help address the knowledge gap that would otherwise undermine the effectiveness of performance-based funding. The importance of feedback in our context is in line with the growing literature that highlights how feedback can contribute to the success of business ventures in low-income countries (e.g., Anderson and McKenzie 2022, De Mel, McKenzie, and Woodruff 2014, Dimitriadis

[10] 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).

and Koning 2022, Giorcelli 2019, McKenzie 2021, McKenzie and Woodruff 2014). Moreover, previous work shows that feedback can be reassuring for individuals and help them build up their self-confidence (Bandiera, Larcinese, and Rasul 2015). This in turn can further reinforce the employees' motivation to tackle the performance targets that are part of the performance-based incentives.

Second, while the above describes how feedback can enhance the effectiveness of performance-based funding, the complementarities between the two likely operate in the other direction as well—that is, performance-based funding can enhance the feedback's effectiveness. Indeed, the previous literature highlights the importance of incentives in promoting learning (e.g., Blimpo 2014, Kremer, Miguel, and Thornton 2009). In this spirit, the combination of feedback with performance-based funding—that is, feedback combined with incentives to act on this feedback and improve health outcomes—is likely to induce the non-profit's staff to pay closer attention to the feedback and build on this feedback to improve the organization's operations.

Third, the complementarities between feedback and performance-based funding are likely to be further enhanced if the feedback is guided by more reliable information—a prerequisite for effective feedback. The reliability of the information that guides the feedback can be improved by further bundling feedback with auditing. The latter is likely to be crucial, as the obtention of relevant and verifiable information is often considered a pervasive concern in low-income countries (e.g., Kerber et al. 2015, Kumar et al. 2022).

In sum, we expect the bundling of performance-based funding and A&F to be an effective tool for non-profit organizations in low-income countries, as the complementarities between the bundle's components reinforce each other in addressing the obstacles faced by non-profits in low-income countries. In this regard, by highlighting the benefits of bundling different governance practices, we echo the insight of Milgrom and Roberts (1988, 1990, 1995) that business practices can be substantially enhanced when complementarities exist between these practices.

In the remainder of this paper, we turn to the empirical analysis to examine the effectiveness of this bundle in the context of a randomized governance program implemented in the DRC.

3. Experimental design

In this section, we describe the PDSS program and the 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 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).[11] 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:

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

[11] As Figure A1 of the Online Appendix shows, the DRC is among the countries with the highest infant mortality rates in the world. In 2017, about 7% of infants died within the first year of their life compared to 0.6% in the U.S. (United Nations 2018).

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

where the first part of the transfer reflects the performance along the 22 healthcare performance indicators listed in Table A1 of the Online Appendix. 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.

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]

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

[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.

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 performance-based incentives combined with A&F—on the health centers' operating efficiency and quality of healthcare services, 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 (i.e., the 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. 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

[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.

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 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). 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 last column) is always insignificant, with *p*-values ranging from 0.130 to 0.981.[16]

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

In Figure 2, we further 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.

[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.

[16] 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. This is further confirmed by the Kolmogorov-Smirnov (KS) test of identical distributions. For the variables plotted in Figure A2, the null of identical distributions cannot be rejected with *p*-values ranging from 0.140 to 0.980.

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 $999 \times 14 = 13{,}986$).

----- 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 3). 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.[17]

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.

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. In Figure 3, we plot the coefficients in event time (along with the corresponding 95% confidence bounds) for each outcome variable.[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 with incentives to actually act on this feedback) were seen as essential in inducing tangible changes that

[17] 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).
[18] In Figure A3, we plot the evolution of each outcome variable separately in the treatment and control groups.
[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.

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 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 specificity of our setting. 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 low-income 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 percentage points. 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 ($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.

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 A8 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., we do not have pre-treatment data).[21]

As can be seen from Table A8, 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 A8 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 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

[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.

number of stillbirths and neonatal deaths documented above.[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 A9 and A10), 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.

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

[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.

strategically refusing high-risk patients (e.g., pregnant mothers who face a higher 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 A11, 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]

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 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

[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.

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 A12 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 A12 (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 A13, 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 group-pair (that is, treated and control health centers are compared to their respective matched outside health centers), we include pair by quarter fixed effects in all regressions. The results are provided in Table 6.

----- Insert Table 6 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 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 18.0% (p-value = 0.000), the number of primary

[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.

healthcare services by 38.6% (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 21.4% (p-value = 0.049)—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 14.8% (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 56.8% (p-value = 0.001), and the number of births by 34.3% (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 operating efficiency and the quality of their healthcare services.

Taken together, the results from Tables 4 and 6 suggest that i) funding alone increases the scale of the health centers but does not improve operating efficiency nor service quality (Table 6); 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 6). As such, these results suggest that funding is not a substitute for governance. Instead, they complement each other in improving the health centers'

overall contribution to health outcomes.

7. Discussion and conclusion

Every year, substantial amounts of taxpayer and donor money are channeled toward improving the healthcare systems of low-income countries. However, an important challenge is to ensure that this funding be used efficiently. This challenge is complicated by the fact that a large share of healthcare services is provided by non-profit health centers that often lack effective governance structures, organizational know-how, and adequate training.

In this paper, we argue that the bundling of performance-based funding with auditing and feedback (A&F) is an effective way to overcome these challenges. The intuition is as follows. First, performance-based funding alone may be ineffective if the health center's managers and employees lack the necessary knowledge to improve the health center's operations. Combining performance-based funding with feedback helps mitigate this challenge. In turn, the bundling of these two provisions—that is, feedback joint with incentives to act on this feedback and achieve specific health outcomes—is likely to enhance the effectiveness of the feedback. Second, the provision of effective feedback requires reliable information—another important challenge in low-income countries—which can be addressed by combining feedback with auditing.

To empirically examine the effectiveness of this bundle, we exploit a unique empirical setting provided by the PDSS program in the DRC. This 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 a governance treatment consisting of performance-based funding combined with A&F. As such, this randomized program provides an ideal setup to study how this bundle of governance practices 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 the bundle of performance-based funding and A&F plays an important role in improving the health centers' operating efficiency and the quality of their services.

In auxiliary analyses, we further examine whether financial subsidies could potentially serve as a substitute for the governance treatment. 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 service quality. This suggests that financial subsidies and the governance treatment operate as complements toward the objective of increasing the health centers' overall contribution to health outcomes.

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 health centers, as well as their donors and impact investors, about the governance mechanisms that are available and effective in achieving their objectives and maximizing the impact of the funds invested.

Our study contributes to several strands of the literature. First, by exploring how a bundle of governance mechanisms consisting of performance-based incentives combined with A&F affects the effectiveness of non-profit health centers in low-income countries, our work contributes to the governance literature and highlights the importance of designing governance practices that take into account the unique nature of non-profits in low-income countries and the obstacles that they face. As such, our study complements 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, Anderson and McKenzie 2022, 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, Iacovone, Maloney, and McKenzie 2022, McKenzie 2021, McKenzie and Woodruff 2014, 2017). Since the non-profit and for-profit contexts are fundamentally different, it is a priori unclear which governance and management practices are effective for non-profits in low-income countries. In particular, while performance-based incentives have been shown to yield substantial efficiency gains in the for-profit sector, the literature that studies the use of performance-based incentives in the non-profit sector of low-income countries finds that such incentives tend to be ineffective, and sometimes even backfire (e.g., Basinga et al. 2011, Eichler and Levine 2009, Huillery and Seban 2021, Linden and Shastry 2012, Miller and Babiarz 2013). A potential explanation is the lack of adequate training and know-how that may undermine these incentives and give rise to unintended consequences (e.g., employees being demotivated). As we argue, the bundling of performance-based incentives with other provisions—namely, regular rounds of feedback (which help mitigate the knowledge gap) and auditing (which helps obtain reliable information that informs the feedback)—is a potential way to overcome these obstacles, which we confirm in our empirical analysis. As such, our study helps move this literature forward by providing new insights on how to think about and design effective performance-based incentives for non-profits in low-income countries.

Second, our paper is related to the literature that emphasizes the importance of complementarities in business practices. In particular, Milgrom and Roberts (1988, 1990, 1995) introduced the "complementarity view" of technological change, arguing that the adoption of new technologies is likely more successful when companies adopt strategic and organizational practices that are complementary to the new technology.[25] Although their context is fundamentally different, our paper builds on the notion of complementarities in an analogous fashion—as we argue, the bundling of performance-based funding and A&F is effective because they complement each other in addressing key challenges faced by non-profits in low-income countries.

Third, our study relates to the large literature that studies the social performance of i) for-profit organizations (e.g., Eccles, Ioannou, and Serafeim 2014, Flammer 2015, Flammer, Hong, and Minor 2019,

[25] As an example, Milgrom and Roberts (1995) describe how the adoption of product line breadth is complementary to technologies that increase manufacturing flexibility.

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 a bundle of governance practices can help non-profits improve health outcomes in low-income countries, and thereby contribute to 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).

Fourth, 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 high-income and emerging countries, much less is known about (for-profit and non-profit) organizations in low-income countries, and especially Africa (Anderson and McKenzie 2022, Dimitriadis and Koning 2022, George et al. 2016b).[26] Yet, understanding how to improve the governance of these organizations—and the implications for operating efficiency and their social impact—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.

[26] For example, the literature on non-profit governance typically focuses on issues that are germane to large non-profits 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).

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, the governance treatment in the PDSS program is administered through a bundle of governance mechanisms (i.e., performance-based funding combined with A&F), and hence we cannot separate between them. Accordingly, another exciting avenue for future research would be to "un-bundle" this governance bundle and quantify the relative importance of each individual provision. Lastly, and relatedly, future research could explore the effectiveness of other types of governance and management practices available to non-profit organizations in low-income countries and track their impact over time.

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.

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.

Anderson SJ, McKenzie D (2022) Improving business practices and the boundary of the entrepreneur: a randomized experiment comparing training, consulting, insourcing, and outsourcing. *J. Political Economy* 130(1):157–209.

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.

Bebchuk LA, Weisbach MS (2010) The state of corporate governance research. *Rev. Financ. Stud.* 23(3):939–961.

Becht M, Bolton P, Roell A (2003) Corporate governance and control. *Handbook of the Economics of Finance* 1(1):1–109.

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.
Blimpo MP (2014) Team incentives for education in developing countries: a randomized field experiment in Benin. *American Econom. J.: Applied Econom.* 6(4):90–109.
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.
Eichler R, Levine R (2009) *Performance Incentives for Global Health: Potential and Pitfalls* (Center for Global Development, Washington, DC).
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.

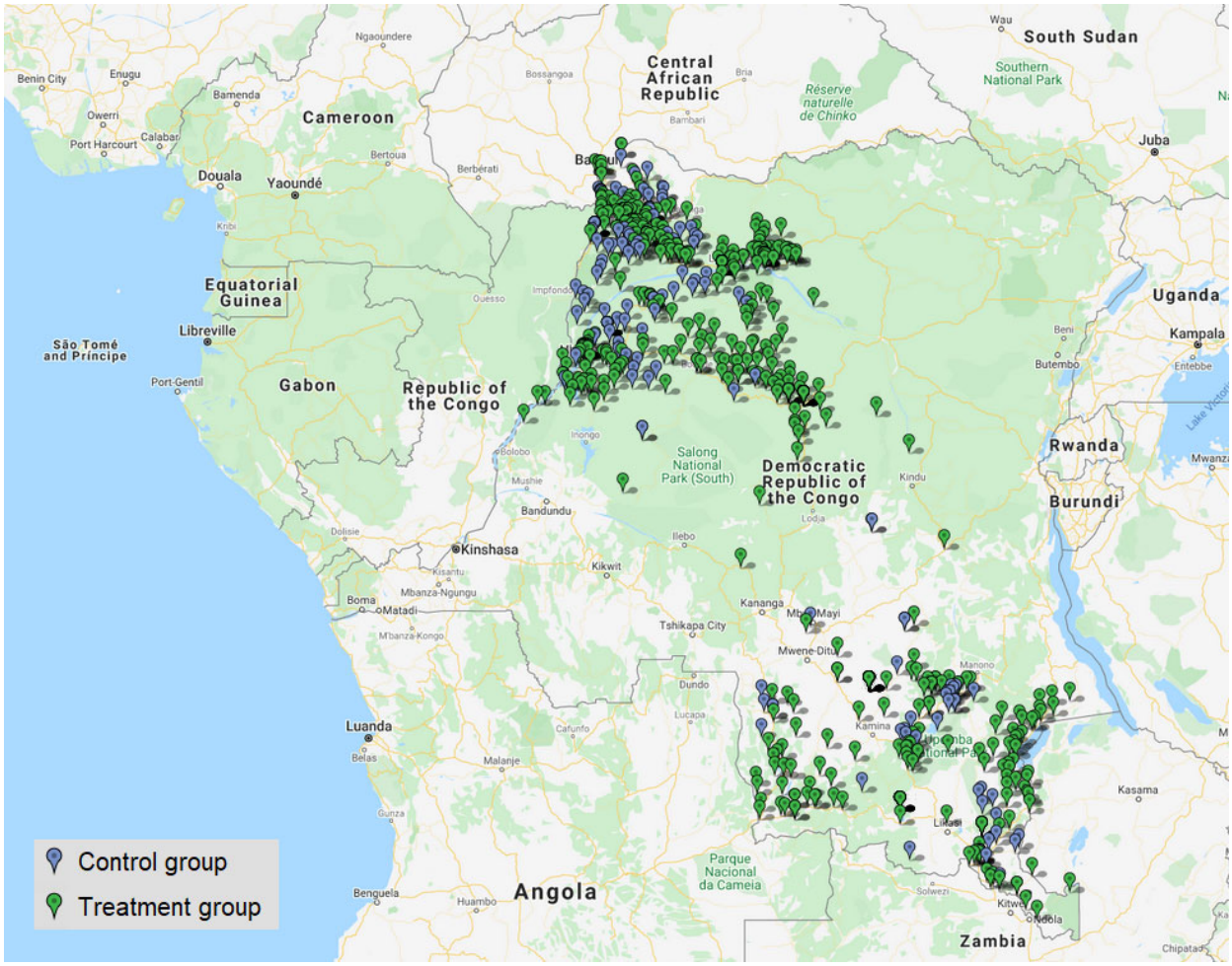
Flammer C, Luo J (2017) Corporate social responsibility as an employee governance tool: evidence from a quasi-experiment. *Strategic Management J.* 38(2):163–183.
Gallus J, Frey BS (2016) Awards: a strategic management perspective. *Strategic Management J.* 37(8):1699–1714.
George G, Howard-Grenville 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.
Glaeser EL (2002) The governance of not-for-profit firms. NBER Working Paper 8921 (National Bureau of Economic Research, Cambridge, MA).
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.
Iacovone L, Maloney W, McKenzie D (2022) Improving management with individual and group-based consulting: results from a randomized experiment in Colombia. *Rev. Econom. Stud.* 89(1):346–371.
IEG (2014) *World Bank Group Support to Health Financing* (World Bank Independent Evaluation Group, Washington, DC).
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.
Kerber KJ, Mathai M, Lewis G, Flenady V, Erwich JJHM, Segun T, Aliganyira P, Abdelmegeid A, Allanson E, Roos N, Rhoda N, Lawn JE, Pattinson R (2015) Counting every stillbirth and neonatal death through mortality audit to improve quality of care for every pregnant woman and her baby. *BMC Pregnancy Childbirth* 15(Suppl. 2):S9.
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.
Kremer M, Miguel E, Thornton R (2009) Incentives to learn. *Rev. Econom. Stat.* 91(3):437–456.
Kumar HNH, Baliga SB, Kushtagi P, Kamath N, Rao SS (2022) Perinatal death auditing is necessary to identify preventable perinatal deaths. *Int. J. Prev. Med.* 13:48.
Leete L (2000) Wage equity and employee motivation in nonprofit and for-profit organizations. *J. Econom. Behav. Organ.* 43(4):423–446.
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.
Milgrom P, Roberts J (1988) Communication and inventories as substitutes in organizing production. *Scandinavian J. Econom.* 90(3):275–289.
Milgrom P, Roberts J (1995) Complementarities and fit strategy, structure, and organizational change in manufacturing. *J. Accounting and Econom.* 19(2–3):179–208.
Milgrom P, Roberts J (1990) The economics of modern manufacturing: technology, strategy, and organization. *American Econom. Rev.* 80(3):511–528.
Miller G, Babiarz KS (2013). Pay-for-performance incentives in low- and middle-income country health programs. NBER Working Paper 18932 (National Bureau of Economic Research, Cambridge, MA).
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.
Paul E, Albert L, N'Sambuka Bisala B, Bodson O, Bonnet E, Bossyns P, Colombo S, De Brouwere V, Dumont A, Sèdjro Eclou D, Gyselinck K, Hane F, Marchal B, Meloni R, Noirhomme M, Noterman JP, Ooms G, Mallé Samb O, Ssengooba F, Touré L, Turcotte-Tremblay AM, Van Belle S, Vinard P, Ridde V (2018) Performance-based financing in low-income and middle-income countries: isn't it time for a rethink? *BMJ Global Health* 3(1):1–7.
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.
Saif-Ur-Rahman KM, Mamun R, Anwar I (2019) Identifying gaps in primary healthcare policy and governance in low-income and middle-income countries: protocol for an evidence gap map. *BMJ Open* 2019(9):1–6.
Scally G, Donaldson LJ (1998) Clinical governance and the drive for quality improvement in the new NHS in England. *BMJ* 317(7150):61–65.
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 (2018) *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 (2021) *WHO's 7 Policy Recommendations on Building Resilient Health Systems* (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

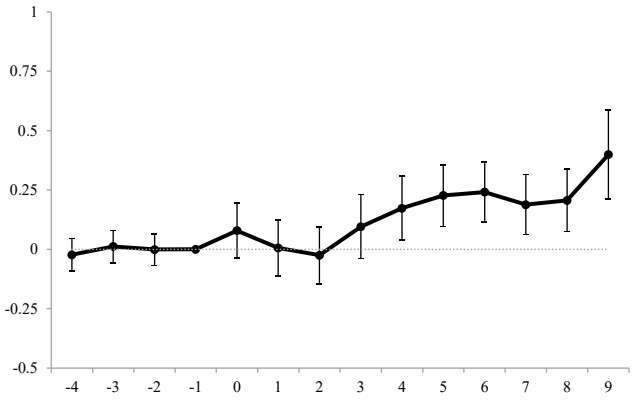


Notes. This figure plots the location of the health centers in the control group (blue markers) and treatment group (green markers), for the health centers in our sample that have non-missing geo-codes (73% of the sample). The map was generated using the online mapping software Maptive.

Figure 3. Dynamics of the treatment effect

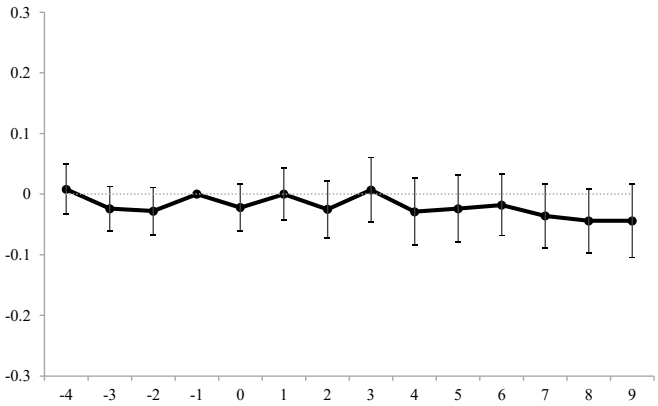
A. Primary healthcare services per employee (log)

Treatment vs. control (with 95% CI)



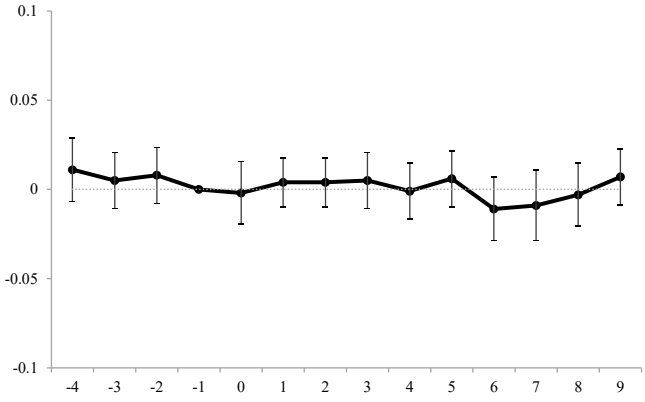
B. Employees (log)

Treatment vs. control (with 95% CI)



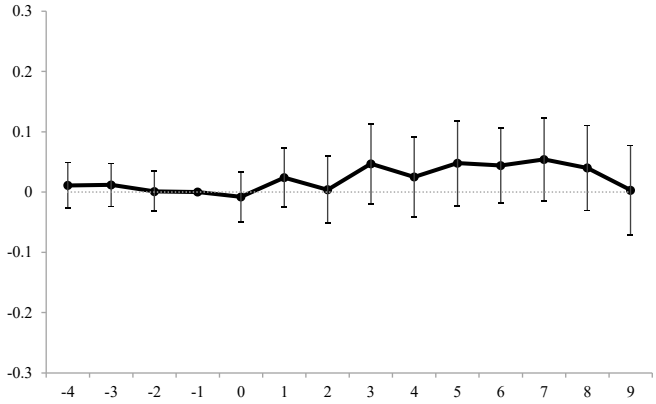
C. Doctors (log)

Treatment vs. control (with 95% CI)



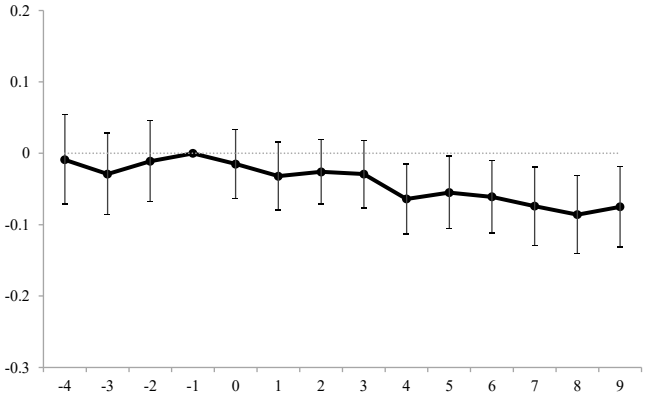
D. Nurses (log)

Treatment vs. control (with 95% CI)



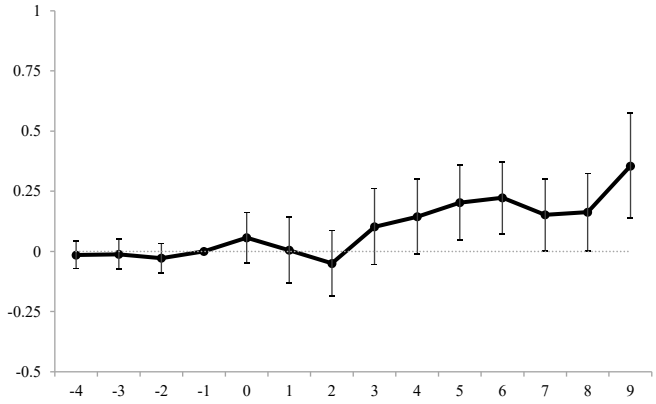
E. Administrative employees (log)

Treatment vs. control (with 95% CI)



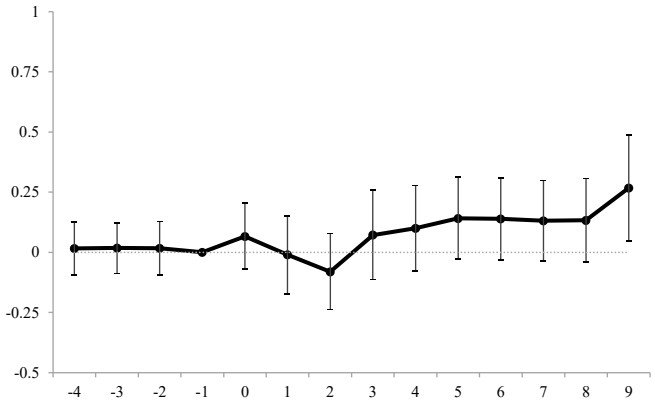
F. Primary healthcare services (log)

Treatment vs. control (with 95% CI)



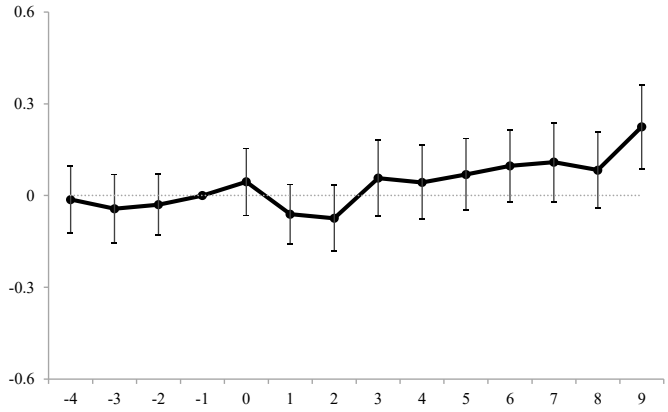
G. Maternal and childhood healthcare services (log)

Treatment vs. control (with 95% CI)



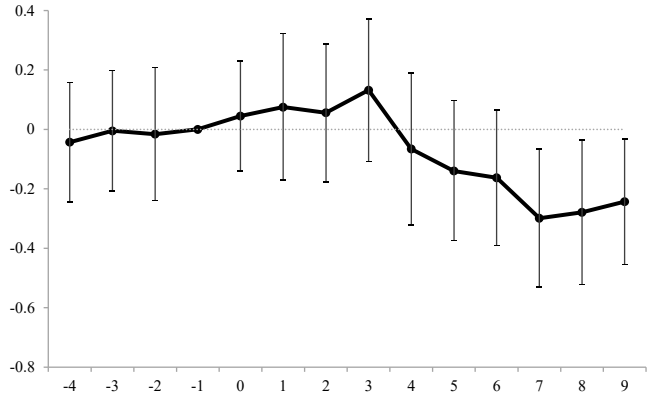
H. Births (log)

Treatment vs. control (with 95% CI)



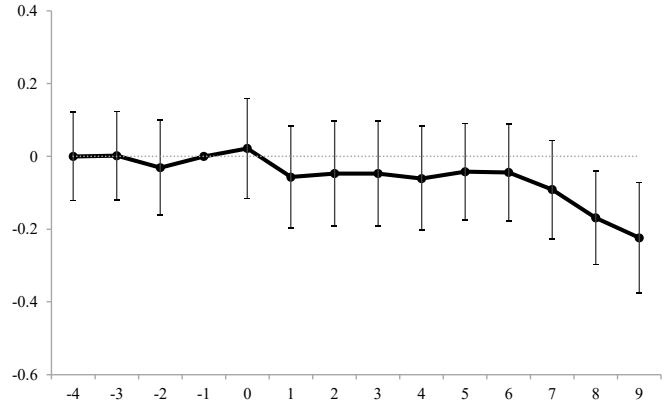
I. Share of stillbirths (%)

Treatment vs. control (with 95% CI)



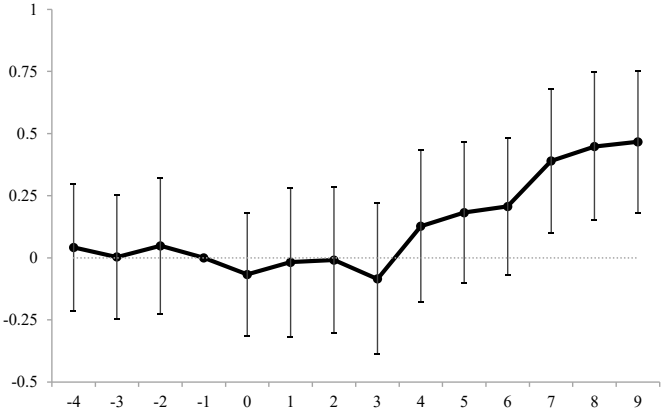
J. Share of neonatal deaths (%)

Treatment vs. control (with 95% CI)



K. Share of live births (%)

Treatment vs. control (with 95% CI)



Notes. This figure plots the coefficients from Table 5 for each outcome variable, along with their corresponding 95% confidence intervals. The *y*-axis refers to the coefficients and their confidence intervals. The *x*-axis indicates quarters in event time, where $t = 0$ is the initial quarter of the PDSS intervention (Q1 2017).

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. Covariate balance

		Obs.	Mean	Median	Std. Dev.	*p*-value (diff. means)
Panel A. Health centers statistics						
Primary healthcare services per employee	Treated	674	238.53	176.62	231.07	0.780
	Control	325	230.02	200.29	173.55	
Employees	Treated	674	7.43	6	7.21	0.196
	Control	325	6.50	6	3.67	
Doctors	Treated	674	0.06	0	0.33	0.457
	Control	325	0.04	0	0.25	
Nurses	Treated	674	3.79	3	4.26	0.130
	Control	325	3.05	3	2.03	
Administrative	Treated	674	3.58	3	3.66	0.710
	Control	325	3.42	3	2.68	
Primary healthcare services	Treated	674	1,660	1,361	1,273	0.351
	Control	325	1,508	1,338	1,009	
Maternal and childhood healthcare services	Treated	674	988	710	1,010	0.494
	Control	325	892	684	770	
Births	Treated	674	55.71	45	41.42	0.348
	Control	325	49.84	45	34.78	
Stillbirths (in %)	Treated	674	0.73	0	1.53	0.500
	Control	325	0.64	0	1.45	
Neonatal deaths (in %)	Treated	674	0.18	0	0.68	0.981
	Control	325	0.17	0	0.68	
Live births (in %)	Treated	674	99.09	100	1.772	0.584
	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
	Control	325	11,227	9,847	9,988	
Population in center's health district	Treated	674	201,829	181,565	73,925	0.716
	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 last column reports the *p*-value of the difference-in-means test 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)	(0.064)	(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. 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)	-0.016	0.039	0.009	0.007	0.035	0.023	0.108	0.045	-0.071	0.015	0.056
	(0.069)	(0.032)	(0.010)	(0.033)	(0.048)	(0.067)	(0.127)	(0.062)	(0.131)	(0.062)	(0.156)
Treatment vs. outside (t = 0 to 4)	0.031	0.045	0.001	0.012	0.008	0.076	0.140	0.072	-0.050	-0.026	0.076
	(0.057)	(0.030)	(0.008)	(0.027)	(0.042)	(0.055)	(0.104)	(0.065)	(0.086)	(0.049)	(0.109)
Control vs. outside (t = 5 to 9)	0.207	0.180***	0.010	0.109**	0.117**	0.386**	0.470**	0.214**	-0.067	-0.080	0.147
	(0.174)	(0.050)	(0.010)	(0.049)	(0.057)	(0.194)	(0.220)	(0.108)	(0.116)	(0.064)	(0.144)
Treatment vs. outside (t = 5 to 9)	0.453***	0.148***	0.002	0.120***	0.046	0.601***	0.568***	0.343***	-0.226***	-0.198***	0.424***
	(0.126)	(0.046)	(0.008)	(0.041)	(0.052)	(0.145)	(0.165)	(0.079)	(0.084)	(0.056)	(0.117)
Health center fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pair by quarter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.527	0.648	0.703	0.643	0.607	0.552	0.553	0.582	0.259	0.241	0.300
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). 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.

ONLINE APPENDIX FOR

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

ANICET A. FANGWA	CAROLINE FLAMMER	MARIEKE HUYSENTRUYT	BERTRAND V. QUÉLIN
HEC Paris	Columbia University and NBER	HEC Paris	HEC Paris
anicet.fangwa@hec.edu	caroline.flammer@columbia.edu	huysentruyt@hec.fr	quelin@hec.fr

CONTENTS

Appendix A. Structure of the DRC's health system

This appendix provides a brief description of the structure of the DRC's health system.[1] The health system of the DRC is organized in three levels (central, provincial, and operational):

- At the *central level* (also referred to as "national level"), the Ministry of Health develops health policies and health standards. It plays a normative role, regulating and supporting the provincial health divisions.
- At the *provincial level* (also referred to as "intermediate level"), each province has a provincial ministry of health with its own provincial health division. Provincial health divisions are mainly responsible for the technical supervision and logistic support of health districts (at the operational level). They also manage the provincial hospital as well as other provincial health structures.
- At the *operational level*, the DRC is divided into 516 health districts (also called "health zones") that cover a population of roughly 100,000 to 200,000. Each health district is administered by a health district management team ("équipe cadre de la zone") that manages a network of health centers and the district hospital. Health districts are further divided into 8,504 health areas. Each health area contains approximately one health center.

Appendix B. Robustness

In this appendix, we describe several robustness checks that are variants of the baseline specifications presented in Table 4.

Specifications with clustered standard errors. In their assessment of the difference-in-differences methodology, Bertrand, Duflo, and Mullainathan (2004) propose several ways to account for serial dependence of the error terms. Two of their recommended methodologies are block-bootstrapping and clustering. In our baseline analysis, we used block-bootstrapped standard errors (defining blocks at the health center level). In Table A5, we re-estimate our baseline regressions using clustered standard errors

[1] For a more detailed description, see WHO (2015).

(defining clusters at the health center level as well). As can be seen, we obtain very similar results as in our baseline analysis.

Multiple hypothesis testing. In Table A6, we supplement our estimates from Table 4 with Anderson's (2008) sharpened *q*-values that account for multiple hypothesis testing. As is shown, our results continue to hold when using this more conservative approach of assessing statistical significance. In particular, the increase in operating efficiency (column (1)) and the increase in the share of live births (column (11)) remain significant the 1% level. Among the other outcome variables, note that the decrease in the number of administrative employees (column (5)) and the decrease in the share of neonatal deaths (column 10)) are now marginally insignificant (both with *q*-value = 0.120). However, when applying Anderson's (2008) sharpened *q*-values to the finer-grained estimates from Table 5, we find that the treatment effect in the last three quarters (that is, at $t = 7$, 8, and 9) remain significant at conventional levels for both variables. For the decrease in the number of administrative staff, the corresponding *q*-values range from 0.044 to 0.085. For the decrease in the share of neonatal deaths, they range from to 0.049 to 0.085.

Weighted least squares. In Table A7, we re-estimate the regressions from columns (9)-(11) of Table 4—that is, the regressions pertaining to the number of stillbirths, the number of neonatal deaths, and the number of live births, respectively—using weighted least squares (WLS), weighting observations by the number of births. Doing so accounts for the fact that ratios of births are more accurately measured when the number of births is higher. As can be seen, we find that our results are slightly stronger when estimated by WLS.

Appendix C. Cross-sectional heterogeneity

The results presented in Section 5 indicate that the governance intervention—i.e., the provision of performance-based incentives combined with A&F—brings about large improvements in operating efficiency and health outcomes. As discussed in the main text, these findings point toward a "learning" interpretation. In particular, the fact that the effects materialize after a few quarters (as opposed to the very

short run) is consistent with the health center staff progressively learning from the quarterly rounds of feedback, and having incentives to act on this feedback.

To shed further light on the mechanism, we examine how our results differ depending on various cross-sectional characteristics. The first variable we consider is the density of health centers in the health district, measured as the number of health centers per capita (i.e., the number of health centers divided by the health district's population). In low-density areas, health centers have limited opportunities to learn and adopt best practices from their peers. Accordingly—to the extent that the auditors' quarterly rounds of feedback help transfer useful knowledge and best practices—the treatment effect might be stronger among those.

The second set of variables we consider are measures of ex ante performance, namely i) operating efficiency and ii) infant mortality (the sum of stillbirths and neonatal deaths, divided by the number of births). Low-performing health centers are more likely to lack proper know-how and hence might be more responsive to the treatment. Finally, we also consider the moderating role of health center size, as smaller health centers might have a lower knowledge base to begin with.

In Table A9, we augment regression (1) by interacting treatment × post (t = 5 to 9) with the four measures described above. Note that all interaction terms are measured in the quarter that precedes the intervention (that is, at $t - 1$), except for infant mortality which is measured over the four quarters prior to the treatment (that is, from $t - 4$ to $t - 1$). We use four quarters to ensure that infant mortality has enough cross-sectional variation, as infant mortality is often 0 for a given health center in a given quarter. For ease of interpretation, we standardize each interaction term. That is, the coefficients capture the differential treatment effect for a one-standard deviation increase in the variable of interest.

As can be seen, we find that the local density of health centers significantly moderates our results. Specifically, we find that health centers in low-density areas experience a higher increase in operating efficiency and a larger decrease in infant mortality. 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. 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. These findings indicate that the improvements brought about by the treatment are larger when there is more to improve and learn to begin with, which again points toward the learning interpretation. In contrast, health center size per se does not significantly moderate the treatment effect.[2]

Naturally, we caution that the results in Table A9 are correlational, as we do not have instruments for the interacted variables.[3] Nevertheless, collectively, these results lend support to the learning interpretation—that is, the health center staff learn from the quarterly rounds of feedback (while being incentivized to act on this feedback), which ultimately translates into higher operating efficiency and improvements in health outcomes.

In Table A10, we refine the analysis of past performance by considering the lower tail of the ex ante performance distribution. Specifically, instead of using continuous measures, we now use two dummy variables that indicate whether the health center is in the bottom 25% and 10%, respectively, of economic performance (measured using operating efficiency; columns (1)-(2)) and service quality (measured using infant mortality; columns (3)-(4)). We find that the treatment effect is significantly stronger among the worse-performing clinics, which confirms that the effectiveness of the intervention is greater among health centers with lower ex ante performance. Interestingly, we find that the treatment effect, albeit smaller in economic terms, remains significant even outside the bottom 25% of clinics based on their ex ante performance. Arguably, given the many challenges faced by the healthcare sector in the DRC, there is considerable room for improvement, even leaving aside the lowest-performing clinics.

[2] The results are similar if the blocks used for the block-bootstrapped standard errors are defined at the health district level (in lieu of the health center level). Doing so accounts for the potential dependence of the error terms along the dimension of the density metric.

[3] In particular, we note that health centers in low-density areas tend to be smaller, have lower operating efficiency, and higher infant mortality. Specifically, density has a correlation of 7.1% (*p*-value = 0.024) with size, 10.2% (*p*-value = 0.001) with operating efficiency, and –7.9% (*p*-value = 0.013) with infant mortality. The inclusion of all four characteristics in the same specification alleviates the possibility that density may merely capture the moderating role of these correlated characteristics.

Appendix D. Potential challenges of the PDSS program and alternative interpretations

D.1 Corruption

A potential challenge is that the health center staff may attempt to bribe the auditors. To alleviate this possibility, the PDSS program has explicit protocols to prevent and sanction fraud. Specifically, the program collaborates with an external counter-audit agency that periodically audits the auditors and helps guarantee the integrity of the process. In case of a discrepancy between the evaluation of the auditors and the evaluation of the counter-audit agency, and if the discrepancy exceeds 10% of the composite score (and no qualitative justification is provided to explain the difference), both the auditor and the health center are sanctioned. For auditors, the first offense triggers the retention of 30% of the auditor team's compensation, and the second offense triggers the suspension of the auditor contract until administrative measures are taken. For health centers, the first penalty is a retention of 30% of the total subsidies from the next quarterly payment, and the second penalty is the termination of the performance contract until administrative measures are taken (such as the replacement of the head of the health center by a candidate whose application is approved by the Ministry of Health).

While these protocols alleviate the possibility of misreporting, it is worth noting that—even if misreporting occurs—it is unlikely to contaminate our analysis. Indeed, our data on health center outcomes are obtained from the DRC Ministry of Health (not the World Bank). Accordingly, we do not examine the specific performance indicators used by the World Bank, but instead study performance outcomes (such as the number of stillbirths, neonatal deaths, etc.) that are collected independently of the PDSS program.

D.2 Employee turnover

Our results show that the treated health centers hire fewer administrative employees post-treatment, while there is no significant change in the hiring of nurses and doctors. That being said, even if we observe no significant change in a given employee category, it could be that low-skill employees were replaced by high-skill employees. In this scenario, our results could be driven by the recruitment of higher-skilled employees, as opposed to the governance treatment itself. Unfortunately, we do not have disaggregated

data at the employee level that could be used to examine this alternative. Nevertheless, we believe it is unlikely to explain our results, for the following reasons.

First, the labor market for health workers is very thin in the DRC. The WHO estimates that the DRC only has 6 qualified health workers per 10,000 population, as compared to the WHO's recommended minimum of 23 per 10,000 population (WHO 2016). Second, and importantly, the mobility of health workers is highly restricted, as their allocation is centralized at the province level. The entry-level market relies on the initial training of medical staff that is exclusively done in the major cities and abroad. Periodically, the National Ministry of Health allocates new health workers to each province, and the respective Provincial Ministry of Health then allocates them to health centers within the province. The reallocation of health workers across health centers is also handled by the respective Provincial Ministry of Health. In other words, each Provincial Ministry of Health has full decision rights over the recruitment and distribution of medical staff across health centers.

Accordingly, labor mobility is not under the direct control of the health centers nor the health workers. What health centers can do is sanction or request the mobility of a specific health worker. The request is then handled at the provincial level. Similarly, health centers can request an increase in their medical staff, but have little influence on the selection process of the personnel they receive. For this reason, it is unlikely that the treated health centers will be able to hire "better" medical staff post-intervention.[4]

D.3 Hawthorne effect

A potential concern with randomized controlled trials is the so-called "Hawthorne effect"—that is, the possibility that the treated subjects might feel observed and alter their behavior in response (for reasons unrelated to the treatment). In particular, it could be that being in the PDSS program induces a status boost that could affect the staff's behavior regardless of the intervention per se. Nevertheless, this concern is

[4] In contrast to the medical staff, non-medical staff (that is, administrative employees) are recruited by the health center board. As such, their number can be optimized at the health center level. Nevertheless, the higher mobility of administrative employees is unlikely to explain the higher quality of medical services per se, as administrative employees are barred from performing medical acts by the PDSS guidelines.

mitigated, for two reasons. First, control health centers were an integral part of the PDSS program, as they received a significant amount of subsidies from the program (the same average amount as the treated health centers). Our qualitative interviews further confirmed that the control health centers clearly felt part of the PDSS program. Second, a pure status boost would likely affect outcomes already in the early quarters of the program, while we find that the effects materialize with a lag of several quarters.

D.4 Implementation challenges

In our interviews of program participants, two implementation challenges were mentioned. First, a non-trivial challenge was accessibility. Some villages and health facilities were harder to access due to poor road conditions, locations in flooded riverine areas, or because of security problems. In several instances, auditors had to use canoes and motorcycles. The second challenge was connectivity. Although auditors were equipped with cell phones and tablets, some teams were out of range of communication systems for days. This limited their ability to take full advantage of the features of the platforms used to enable online data collection and retrieval. These logistical challenges led to occasional delays in conducting the auditing. That being said, given their idiosyncratic nature, they are unlikely to substantially affect our analysis.[5]

D.5 "Superstar" health centers

In principle, the intervention could lead to a separation of the treated health centers into "superstar" health centers (i.e., those that do well based on the performance targets) and "failure" health centers (i.e., those that do not). In this scenario, the tails of the performance improvements would be wider in the treatment vs. control group. In Figure A5, we explore this possibility by plotting the post-treatment distributions of each outcome variable in the treatment and control groups, respectively.

In Panel A, we plot the post-treatment distributions of operating efficiency. As can be seen, we find that the right tail improves in the treatment (vs. control) group. However, we do not find evidence for a

[5] Another potential challenge was the Ebola outbreak of 2014-2016 that preceded the launch of the PDSS program. While the Ebola crisis most certainly affected the health centers in our sample, it is unlikely to significantly distort our estimates as it affected both the treatment and control health centers.

deterioration in the left tail. In fact, the left tail improves as well. More broadly, the whole distribution seems to shift to the right (compared to the control group). In Panel K, we observe a similar pattern when looking at the post-treatment distributions of the share of live births. (A consistent pattern is also observed in Panels I and J when looking at the share of stillbirths and share of neonatal deaths, respectively.) Overall, this suggests that the treated health centers that score low in the quarterly assessments (and hence receive less funding than the control health centers) do not perform worse than the control group.

This finding is consistent with the World Bank's insights from their earlier pilot programs in which they found that the sole provision of performance-based funding did not affect the performance of health centers in the RDC. This is also consistent with our auxiliary findings based on the "outside" group, where we find that funding per se is not associated with subsequent changes in performance (see Section 6). Rather, this finding highlights the importance of combining performance-based incentives with quarterly rounds of auditing and feedback. By implementing at least some of the feedback—which is likely to be the case for most treated health centers—performance improves. In contrast, not implementing the feedback is conducive to the status quo and hence need not hurt performance (compared to the counterfactual provided by the control group).

References in online appendix

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.

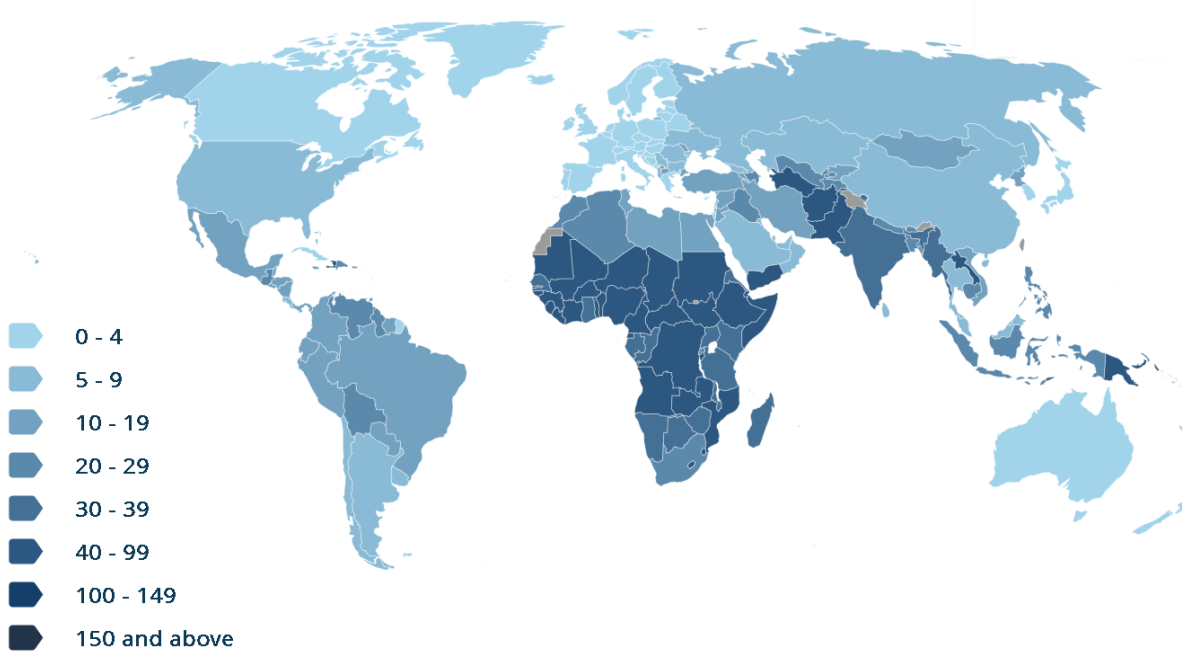
Bertrand M, Duflo E, Mullainathan S (2004) How much should we trust differences-in-differences estimates? *Quart. J. Econom.* 119(1):249–275.

PDSS (2016) *Manuel Opérationnel du Financement Basé sur la Performance* (Ministry of Health, Democratic Republic of Congo).

WHO (2015) *Democratic Republic of the Congo: Improving Aid Coordination in the Health Sector* (World Health Organization, Geneva).

WHO (2016) *Achieving the Health-Related MDGs. It Takes a Workforce!* (World Health Organization, Geneva).

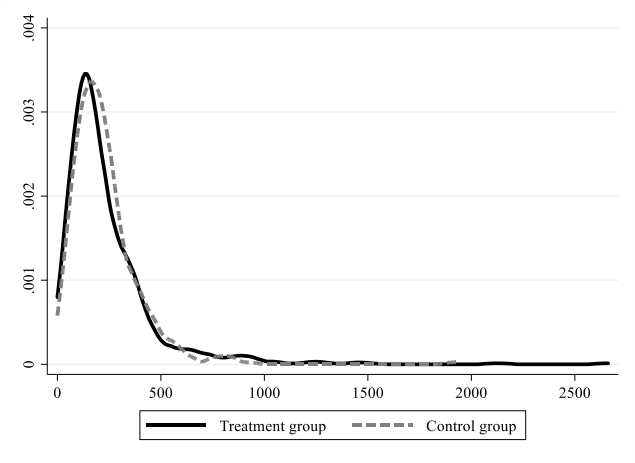
Figure A1. Infant mortality rate



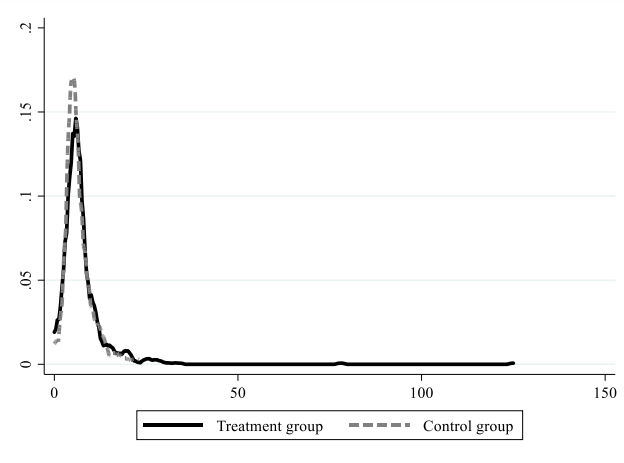
Notes. Infant mortality rate is measured by the number of deaths of infants under one-year old per 1,000 births alive. Source: United Nations Inter-Agency Group for Child Mortality Estimation (https://childmortality.org/data).

Figure A2. Pre-treatment distributions

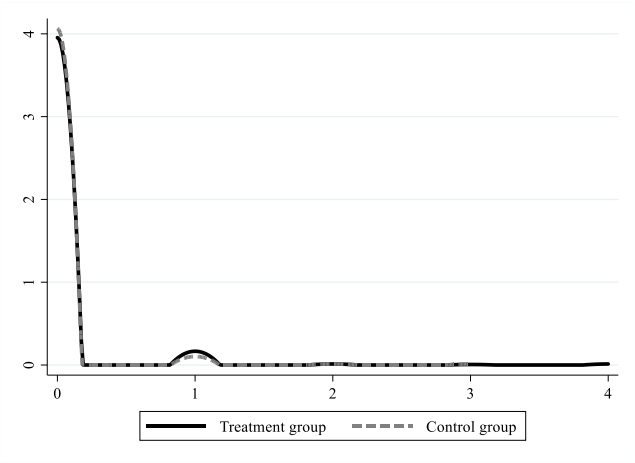
A. Primary healthcare services per employee at $t - 1$



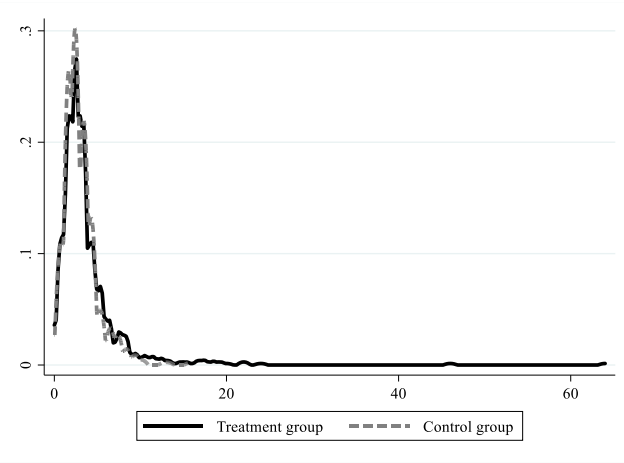
B. Employee at $t - 1$



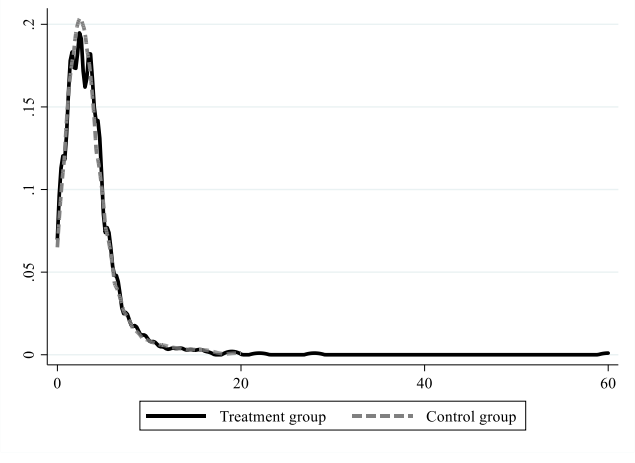
C. Doctors at $t - 1$



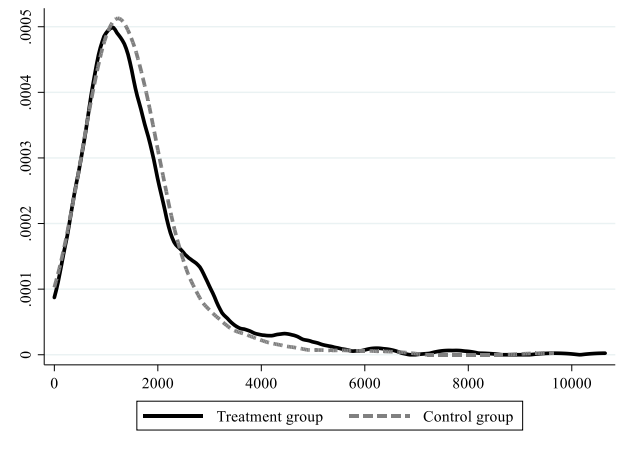
D. Nurses at $t - 1$



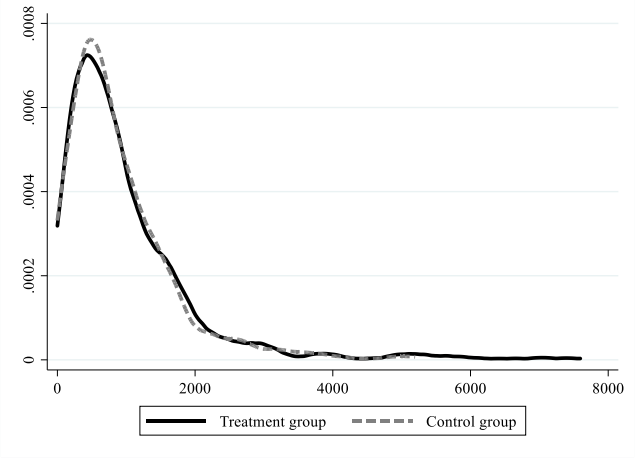
E. Administrative employees at $t - 1$



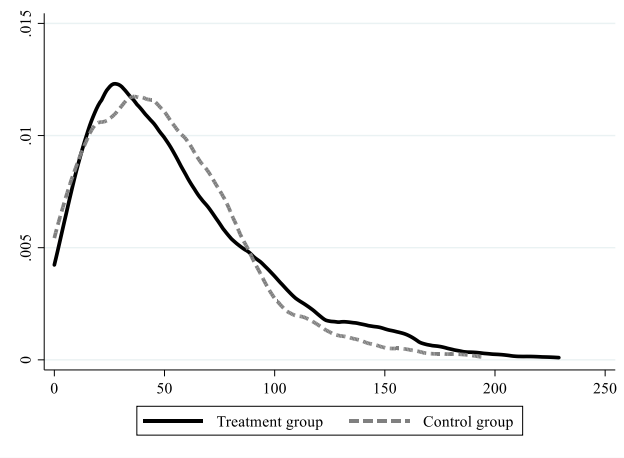
F. Primary healthcare services at $t - 1$



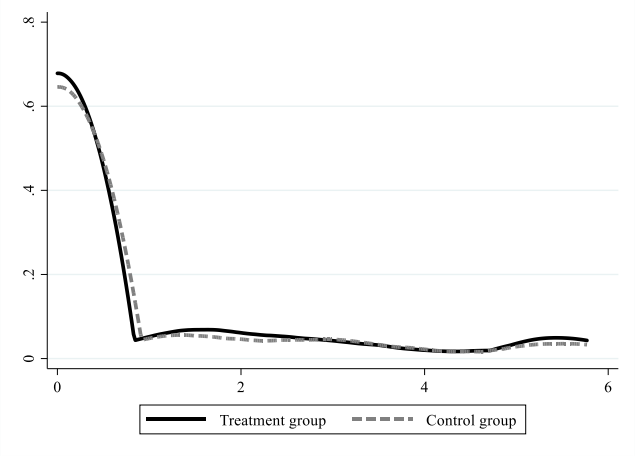
G. Maternal and childhood healthcare services at $t - 1$



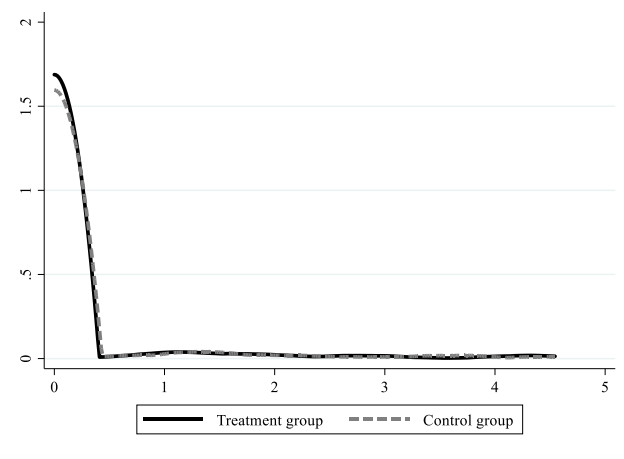
H. Births at $t - 1$



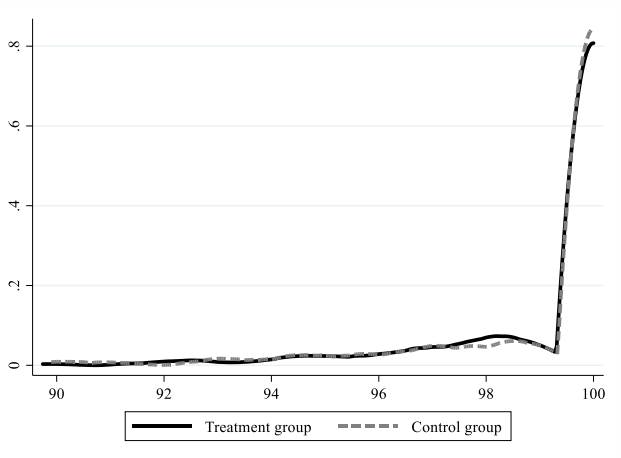
I. Share of stillbirths at $t - 1$



J. Share of neonatal deaths at $t - 1$

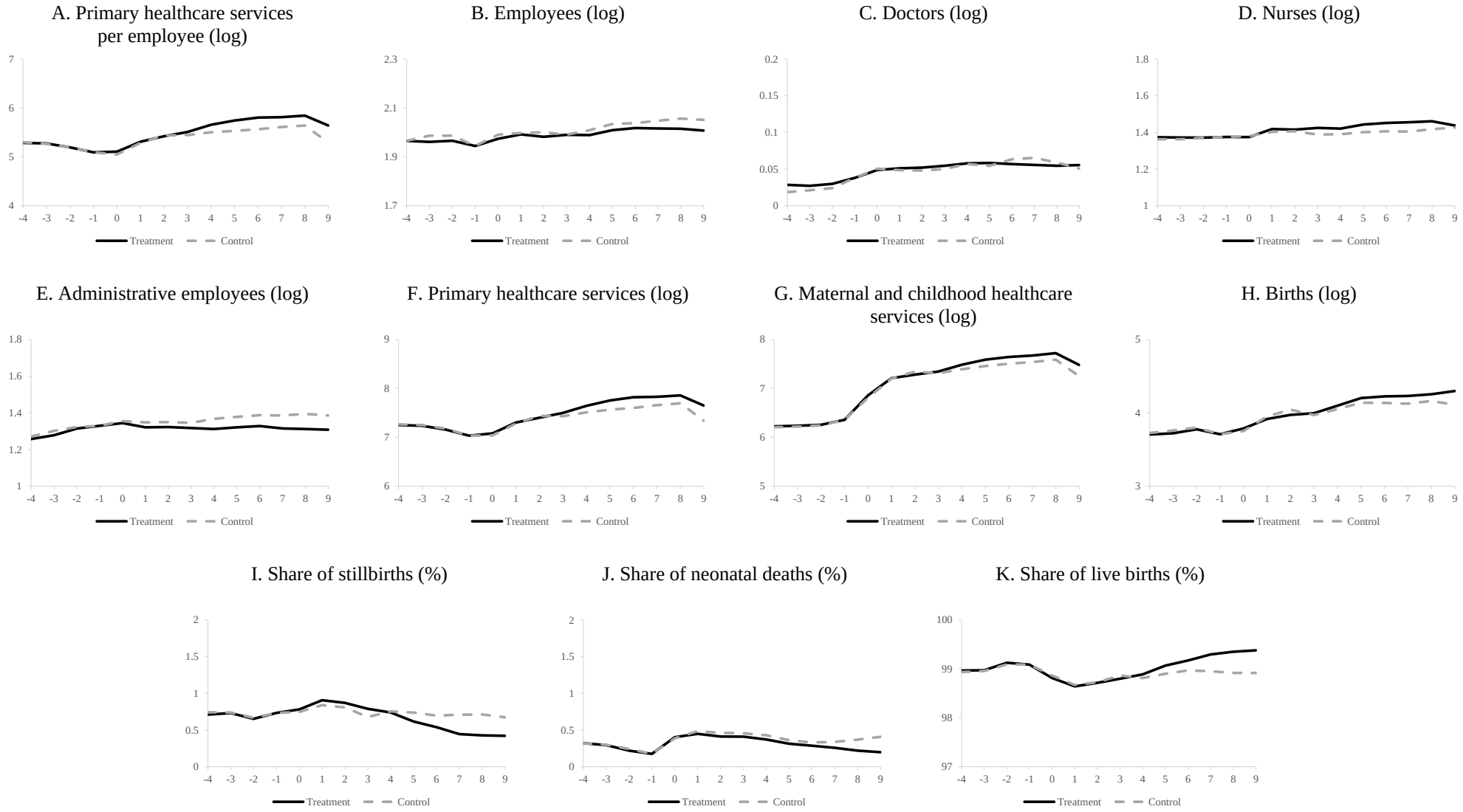


K. Share of live births at $t - 1$



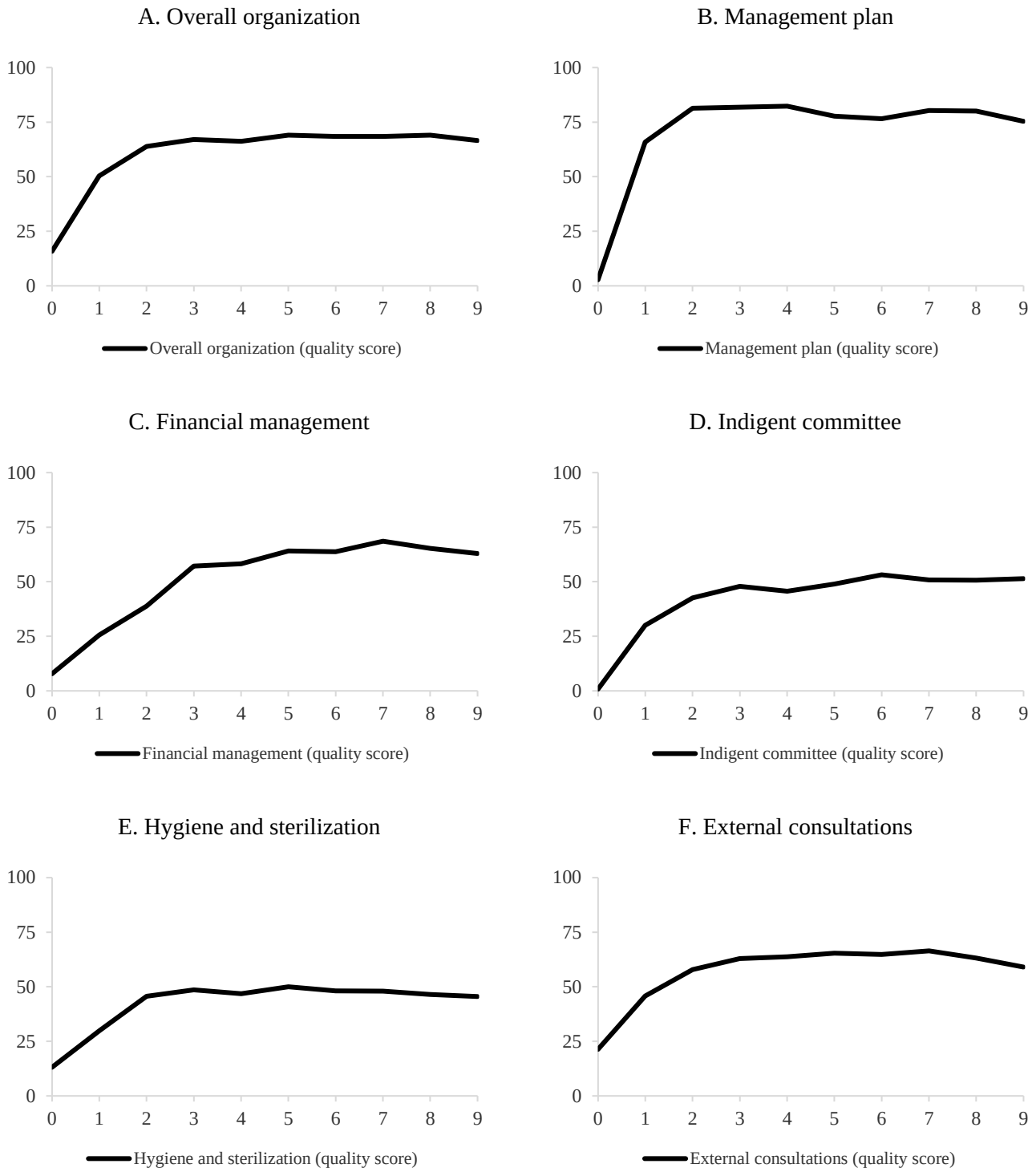
Notes. This figure plots the kernel density of each outcome variable at $t - 1$ (Q4 2016), that is, in the quarter that precedes the PDSS intervention, separately for the treatment (black solid line) and control (gray dashed line) groups.

Figure A3. Dynamics

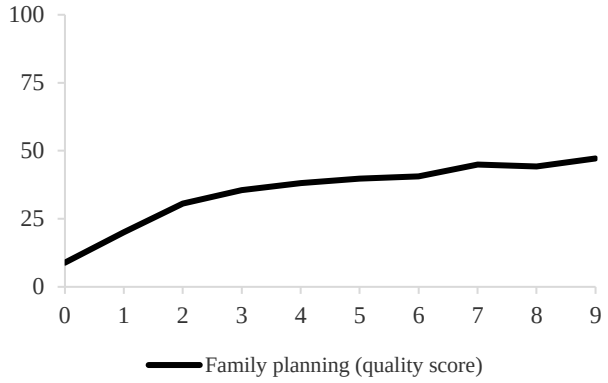


Notes. This figure plots the evolution of each outcome variable in the treatment and control groups. In each panel, the *y*-axis refers to the mean of the respective outcome variable across all health centers in the treatment (black solid line) and control (gray dashed line) groups, respectively. The *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 $t = -1$.

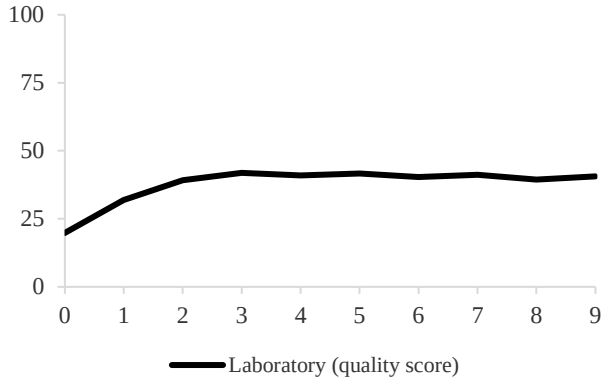
Figure A4. Quality improvements at the treated health centers



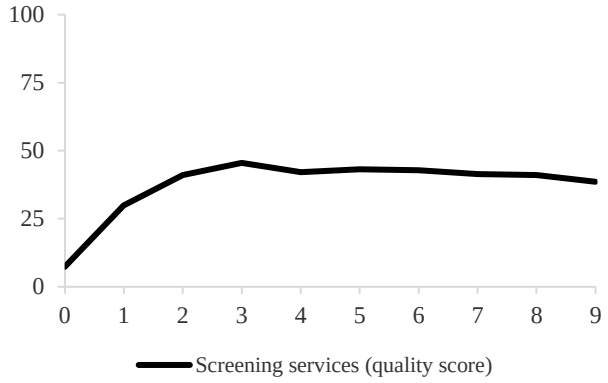
G. Family planning



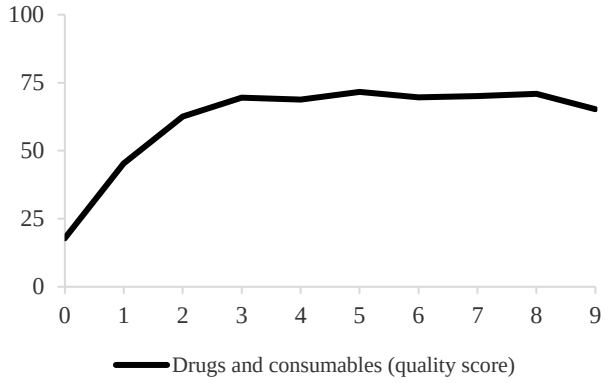
H. Laboratory



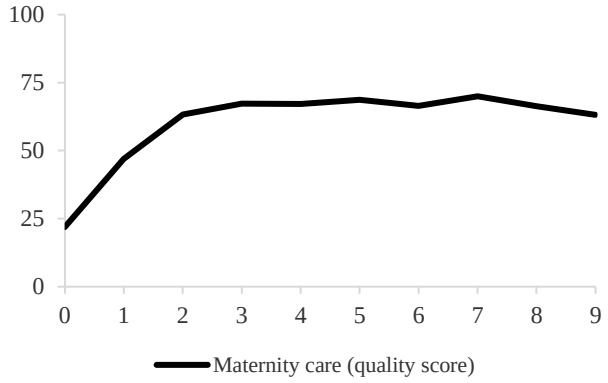
I. Screening services



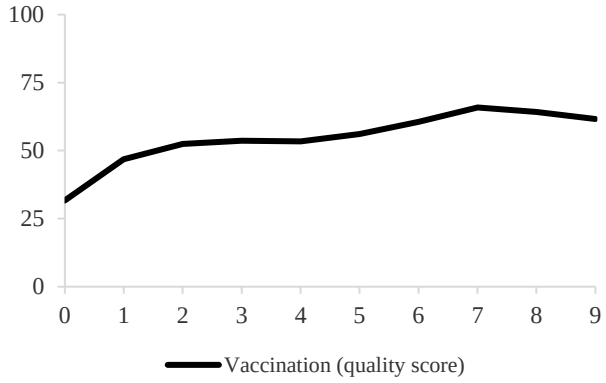
J. Drugs and consumables

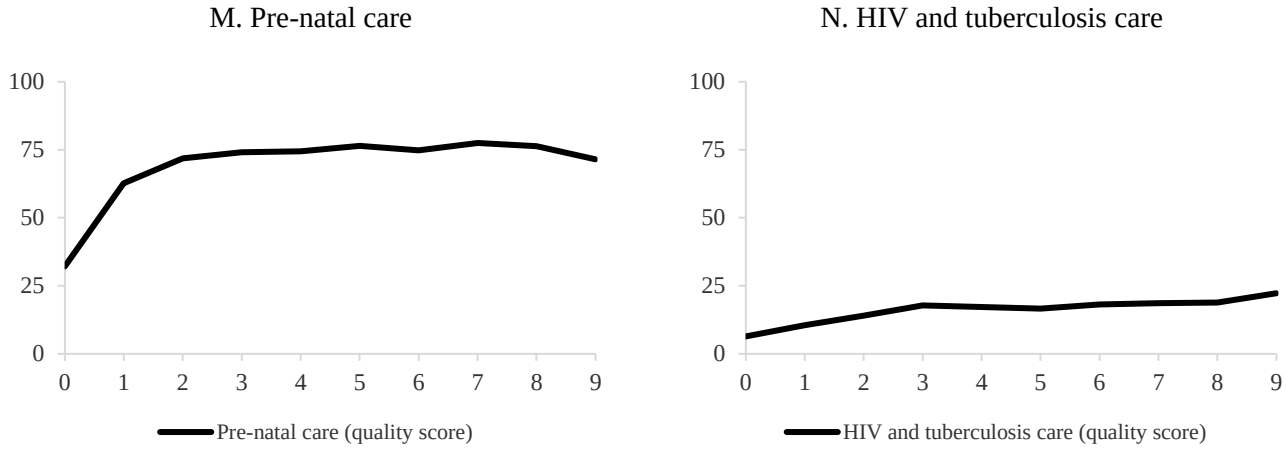


K. Maternity care



L. Vaccination

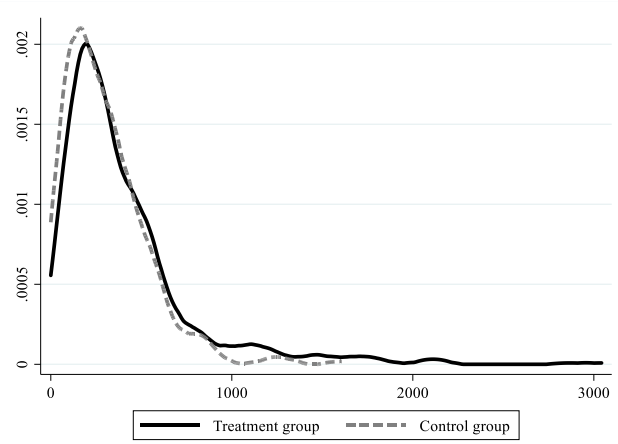




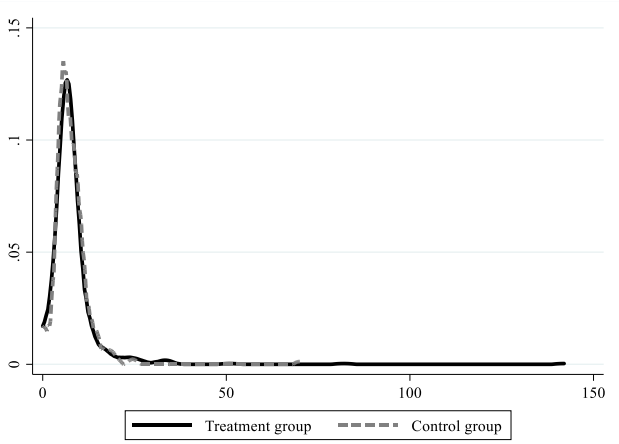
Notes. This figure plots the mean of the treated health centers' quality scores on a quarterly basis from $t = 0$ (Q1 2017) until $t = 9$ (Q2 2019). In each panel, the *y*-axis refers to the mean of the respective quality score. The *x*-axis indicates quarters in event time, where $t = 0$ is the initial quarter of the PDSS intervention (Q1 2017). The quality scores are described in Table A3. The sample consists of 646 treated health centers with non-missing quality scores.

Figure A5. Post-treatment distributions

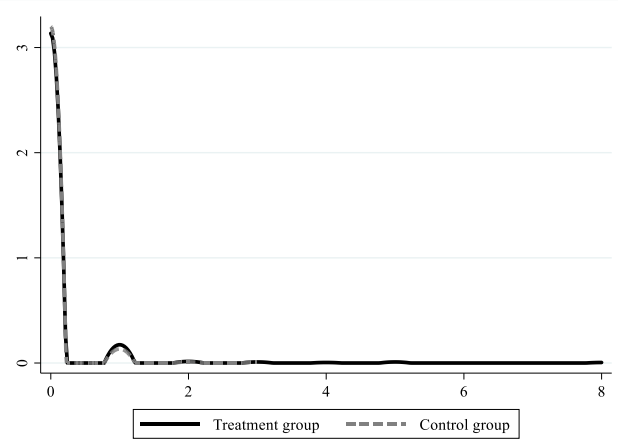
A. Primary healthcare services per employee at $t + 9$



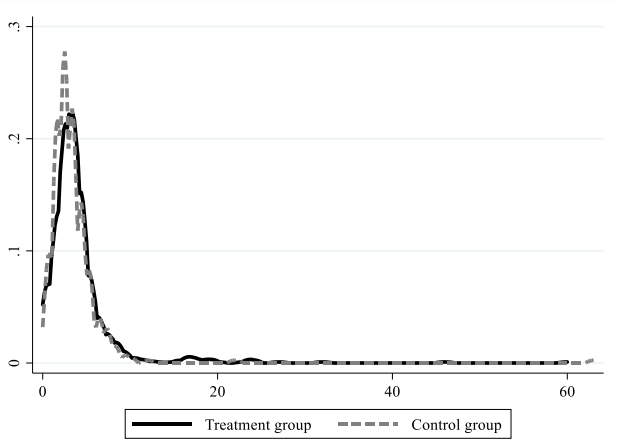
B. Employee at $t + 9$



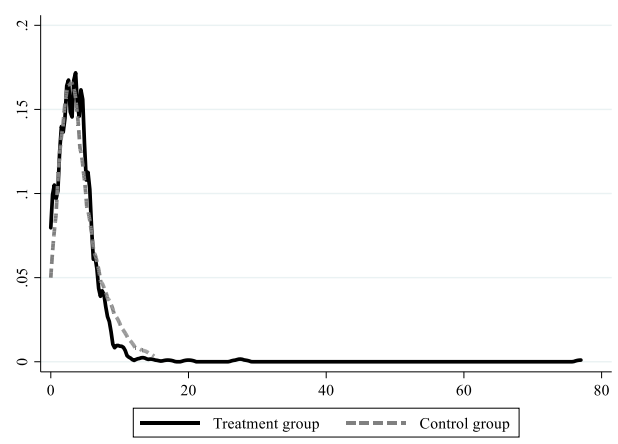
C. Doctors at $t + 9$



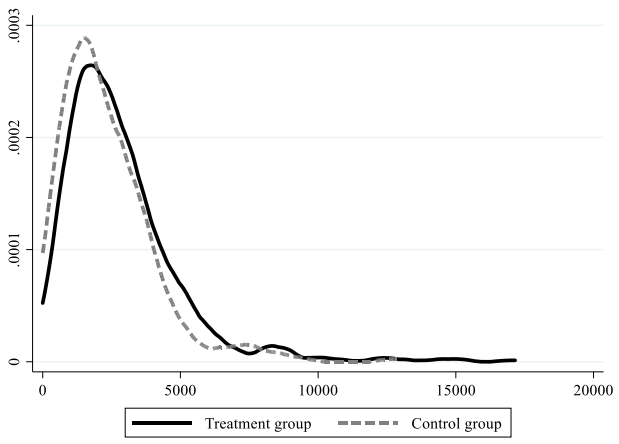
D. Nurses at $t + 9$



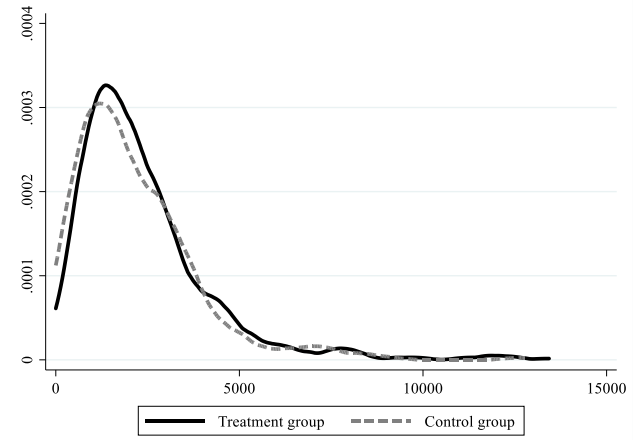
E. Administrative employees at $t + 9$



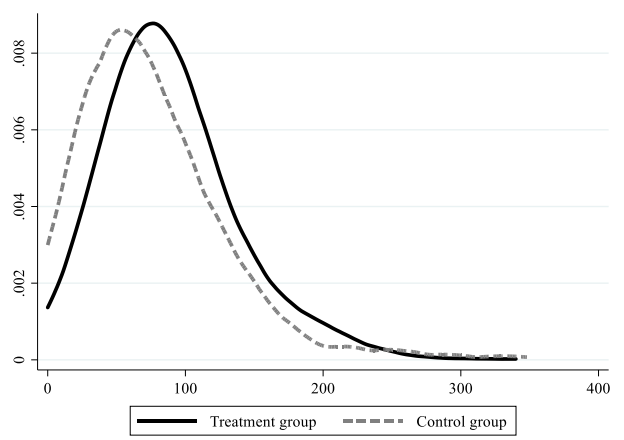
F. Primary healthcare services at $t + 9$



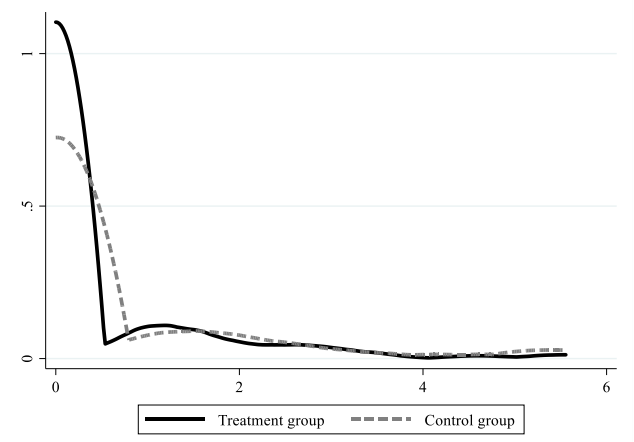
G. Maternal and childhood healthcare services at $t + 9$



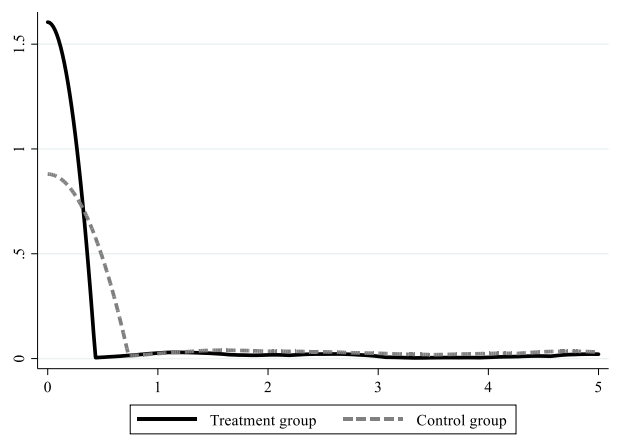
H. Births at $t + 9$



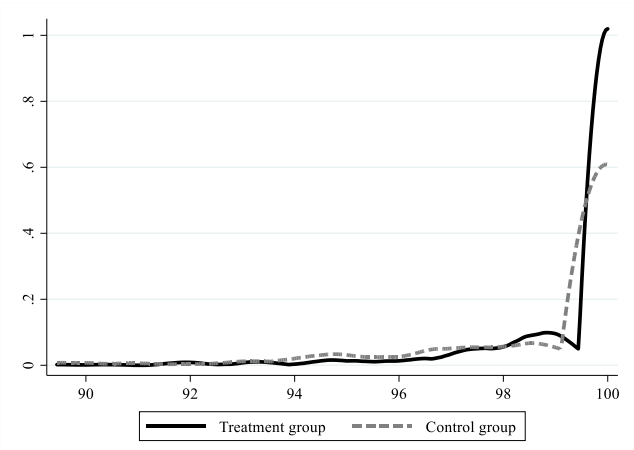
I. Share of stillbirths at $t + 9$



J. Share of neonatal deaths at $t + 9$



K. Share of live births at $t + 9$



Notes. This figure plots the kernel density of each outcome variable at $t + 9$ (Q2 2019), that is, ten quarters after the PDSS intervention, separately for the treatment (black solid line) and control (gray dashed line) groups.

Table A1. Output indicators

1	Number of new curative care visits
2	Number of new curative care visits (vulnerable groups)
3	Number of minor surgeries
4	Number of severe cases referred to hospitals
5	Number of children immunized
6	Number of women receiving at least two doses of tetanus toxoid vaccine during prenatal care
7	Number of women receiving three doses of sulfadoxine-pyrimethamine during prenatal care
8	Number of first prenatal care visits
9	Number of fourth prenatal care visits
10	Number of second and third postnatal care visits
11	Number of assisted deliveries
12	Number of first-time family planning visits pertaining to short-acting contraceptive methods (injectable or oral)
13	Number of first-time family planning visits pertaining to long-term contraceptive methods (IUD or implants)
14	Number of growth-monitoring visits for children aged between 6 and 23 months (preventive care)
15	Number of growth-monitoring visits for children aged between 24 and 59 months (preventive care)
16	Number of home visits
17	Number of voluntary HIV counseling and testing visits
18	Number of HIV+ pregant women receiving antiretroviral therapy
19	Number of HIV+ newborns receiving antiretroviral therapy
20	Number of HIV+ patients receiving antiretroviral therapy
21	Number of patients tested for pulmonary tuberculosis
22	Number of patients cured from pulmonary tuberculosis

Source: PDSS (2016, pp. 60-61).

Table A2. Weights for the quality score

		Maximum quality score	Weight
1	Overall organization	31	7.8%
2	Management plan	9	2.3%
3	Financial management	15	3.8%
4	Indigent committee	20	5.0%
5	Hygiene and sterilization	31	7.8%
6	External consultations	128	32.0%
7	Family planning	32	8.0%
8	Laboratory	17	4.3%
9	Screening services	6	1.5%
10	Drugs and consumables	25	6.3%
11	Tracer drugs	17	4.3%
12	Maternity care	24	6.0%
13	Vaccination	20	5.0%
14	Pre-natal care	12	3.0%
15	HIV and tuberculosis care	13	3.3%
	Total	400	100.0%

Source: PDSS (2016, p. 92).

Table A3. Description of the quality criteria

1	Overall organization	Monthly technical meetings; home visits (team composition, follow-up); documentation of meetings; organigram.
2	Management plan	Evaluation of process (how management plan was construed); content of management plan; analysis of other health centers in the areas; subcontracting strategy; mobile health strategies.
3	Financial management	Documents exist; bonus pay based on performance is known and exists.
4	Indigent committee	Monthly meetings of committee of vulnerables; budget is set aside for taking care of vulnerables.
5	Hygiene and sterilization	Suitable infrastructure exists and is organized in a certain way (e.g., garbage is at least 2 meters deep and 15 meters away from the center); recycling of medical waste.
6	External consultations	Infrastructure of waiting room; transparency about the prices charged; electricity; staff is wearing appropriate clothes; review of protocols applied to treat severe malaria; availability of medical equipment; evaluation of the health status of children below 5 years of age.
7	Family planning	Availability of infrastructure and devices; trained personnel; medical fiches are filled in.
8	Laboratory	Technician is available; fiches are filled in correctly; recycling and cleaning of material.
9	Screening services	Infrastructure for both personnel and patients; fiches are filled in.
10	Drugs and consumables	Stock of medicines exists and is correctly organized; distribution of drugs (generic, only for medical purposes).
11	Tracer drugs	Minimal threshold of specific drugs available is met.
12	Maternity care	Infrastructure; medical equipment; delivery bed; newborn care.
13	Vaccination	Vaccination rate; vaccines availability and storage.
14	Pre-natal care	Scale (infrastructure); personnel; fiches.
15	HIV and tuberculosis care	Scale (infrastructure); personnel; fiches.

Source: PDSS (2016, pp. 220-256).

Table A4. Minimum detectable effect sizes

	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)
Coefficient (from Table 4)	0.255	-0.022	-0.008	0.032	-0.058	0.233	0.149	0.138	-0.209	-0.107	0.316
MDES	0.166	0.059	0.026	0.101	0.080	0.190	0.215	0.134	0.200	0.148	0.276
| Coefficient | > MDES?	Yes	No	No	No	No	Yes	No	Yes	Yes	No	Yes

Notes. This table reports the minimum detectable effect size (MDES) pertaining to the coefficient of treatment × post ($t = 5$ to $t = 9$) in Table 4 for each outcome variable. The MDES is computed using a 5% significance level and 80% power.

Table A5. Robustness—clustered standard errors

	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)	(0.017)	(0.008)	(0.028)	(0.024)	(0.045)	(0.062)	(0.041)	(0.070)	(0.050)	(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***
	(0.060)	(0.021)	(0.009)	(0.036)	(0.028)	(0.068)	(0.075)	(0.050)	(0.074)	(0.051)	(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. This table reports variants of the regressions in Table 4, using clustered (in lieu of block-bootstrapped) standard errors. Standard errors are clustered at the health center level. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.

Table A6. Robustness—Anderson's (2008) sharpened *q*-values

	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
	[0.201]	[0.879]	[0.626]	[0.705]	[0.453]	[0.248]	[0.817]	[0.626]	[0.453]	[0.626]	[0.770]
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***
	[0.001]	[0.453]	[0.453]	[0.453]	[0.120]	[0.007]	[0.126]	[0.016]	[0.016]	[0.120]	[0.010]
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. This table provides the sharpened *q*-values of Anderson (2008) pertaining to the estimates from Table 4. The *q*-values are reported in squared brackets. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively.

Table A7. Robustness—birth-weighted regressions

	Share of stillbirths (%)	Share of neonatal deaths (%)	Share of live births (%)
	(1)	(2)	(3)
Treatment × post (t = 0 to 4)	0.060	-0.026	-0.035
	(0.070)	(0.055)	(0.095)
Treatment × post (t = 5 to 9)	-0.247***	-0.155***	0.402***
	(0.074)	(0.060)	(0.110)
Health center fixed effects	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes
R-squared	0.346	0.370	0.416
Observations	13,986	13,986	13,986

Notes. This table provides variants of the regressions in columns (9)-(11) of Table 4, but estimating the regressions by weighted least squares (WLS), weighting observations by the number of births. 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 A8. 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 A9. Cross-sectional heterogeneity

	Health center operating efficiency	Quality of healthcare services		
	Primary healthcare services per employee (log)	Share of stillbirths (%)	Share of neonatal deaths (%)	Share of live births (%)
	(1)	(2)	(3)	(4)
Treatment × post (t = 0 to 4)	0.068	0.065	-0.031	-0.034
	(0.042)	(0.069)	(0.052)	(0.095)
Treatment × post (t = 5 to 9)	0.242***	-0.200***	-0.105**	0.305***
	(0.052)	(0.066)	(0.052)	(0.092)
Treatment × post (t = 5 to 9) × Density of health centers	-0.105***	0.071**	0.055	-0.126**
	(0.034)	(0.033)	(0.038)	(0.059)
Treatment × post (t = 5 to 9) × Infant mortality	-0.027	-0.493***	-0.118***	0.611***
	(0.025)	(0.039)	(0.040)	(0.066)
Treatment × post (t = 5 to 9) × Size	-0.048	0.044	-0.011	-0.032
	(0.033)	(0.026)	(0.024)	(0.041)
Treatment × post (t = 5 to 9) × Operating efficiency	-0.497***	-0.003	-0.038*	0.041
	(0.045)	(0.023)	(0.021)	(0.036)
Health center fixed effects	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes
R-squared	0.599	0.309	0.327	0.373
Observations	13,986	13,986	13,986	13,986

Notes. This table presents variants of the regressions in columns (1) and (9)-(11) of Table 4, interacting treatment × post (t = 5 to t = 9) with various characteristics measured prior to the PDSS intervention. For ease of interpretation, all interacted characteristics are standardized. 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 A10. Low-performing health centers

	Health center operating efficiency		Quality of healthcare services	
	Primary healthcare services per employee (log)	Primary healthcare services per employee (log)	Share of live births (%)	Share of live births (%)
	(1)	(2)	(3)	(4)
Treatment × post (t = 0 to 4)	0.068	0.068	-0.034	-0.034
	(0.042)	(0.042)	(0.095)	(0.095)
Treatment × post (t = 5 to 9)	0.113**	0.159***	0.181*	0.237**
	(0.052)	(0.052)	(0.096)	(0.097)
Treatment × post (t = 5 to 9) × Operating efficiency (bottom 25%)	0.565***			
	(0.098)			
Treatment × post (t = 5 to 9) × Operating efficiency (bottom 10%)		0.919***		
		(0.140)		
Treatment × post (t = 5 to 9) × Infant mortality (bottom 25%)			0.589***	
			(0.133)	
Treatment × post (t = 5 to 9) × Infant mortality (bottom 10%)				0.749***
				(0.219)
Health center fixed effects	Yes	Yes	Yes	Yes
Quarter fixed effects	Yes	Yes	Yes	Yes
R-squared	0.569	0.576	0.358	0.358
Observations	13,986	13,986	13,986	13,986

Notes. This table presents variants of the regressions in columns (1) and (11) of Table 4, interacting treatment × post (t = 5 to t = 9) with two dummy variables that indicate whether operating efficiency and infant mortality, respectively, are in the bottom quartile ("bottom 25%") and bottom decile ("bottom 10%") of their distribution prior to 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 A11. 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 A12. Summary statistics for the outside group

	Obs.	Mean	Median	Std. Dev.
Panel A. Health centers statistics				
Primary healthcare services per employee	5,832	168.10	80.07	429.14
Employees	5,832	5.20	4	6.64
Doctors	5,832	0.23	0	0.86
Nurses	5,832	2.82	2	4.69
Administrative	5,832	2.15	1	2.59
Primary healthcare services	5,832	1,153	421	2,843
Maternal and childhood healthcare services	5,832	607	262	979
Births	5,832	26.77	13	41.39
Stillbirths (in %)	5,832	0.41	0	1.30
Neonatal deaths (in %)	5,832	0.37	0	1.95
Live births (in %)	5,832	99.22	100	2.55
Panel B. Population statistics				
Population in center's health area	5,832	11,918	9,495	8,700
Population in center's health district	5,832	215,158	194,315	94,966

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

Table A13. Covariate balance post matching

	Treatment vs. matched outside group		Control vs. matched outside group	
	Obs.	*p*-value (diff. in means)	Obs.	*p*-value (diff. in means)
	(1)	(2)	(3)	(4)
Primary healthcare services per employee	1,348	0.342	650	0.516
Employees	1,348	0.659	650	0.638
Doctors	1,348	0.402	650	0.485
Nurses	1,348	0.433	650	0.822
Administrative	1,348	0.950	650	0.512
Primary healthcare services	1,348	0.878	650	0.820
Maternal and childhood healthcare services	1,348	0.861	650	0.622
Births	1,348	0.414	650	0.649
Stillbirths (in %)	1,348	0.518	650	0.343
Neonatal deaths (in %)	1,348	0.420	650	0.790
Live births (in %)	1,348	0.822	650	0.375

Notes. All variables are recorded in the quarter preceding the PDSS intervention (that is, Q4 2016). The table reports the *p*-value of the difference-in-means test comparing treated health centers vs. matched outside health centers (columns (1) and (2)), and control health centers vs. matched outside health centers (columns (3) and (4)). 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.