

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
HEC Paris
anicet.fangwa@hec.edu

CAROLINE
FLAMMER
Columbia University and NBER
caroline.flammer@columbia.edu

MARIEKE
HUYSENTRUYT
HEC Paris
huystruyt@hec.fr

BERTRAND V.
QUÉLIN
HEC Paris
quelin@hec.fr

CONTENTS

Appendix A. Structure of the DRC's health system

Appendix B. Robustness

Appendix C. Cross-sectional heterogeneity

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

Appendix figures and tables

Figure A1. Infant mortality rate

Figure A2. Pre-treatment distributions

Figure A3. Dynamics

Figure A4. Quality improvements at the treated health centers

Figure A5. Post-treatment distributions

Table A1. Output indicators

Table A2. Weights for the quality score

Table A3. Description of the quality criteria

Table A4. Minimum detectable effect sizes

Table A5. Robustness—clustered standard errors

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

Table A7. Robustness—birth-weighted regressions

Table A8. Quality improvements at the treated health centers

Table A9. Cross-sectional heterogeneity

Table A10. Low-performing health centers

Table A11. Quality of healthcare services at other healthcare facilities in the same health district
as the treated health centers

Table A12. Summary statistics for the outside group

Table A13. Covariate balance post matching

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 $t = 7, 8, \text{ and } 9$) 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 \times 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.

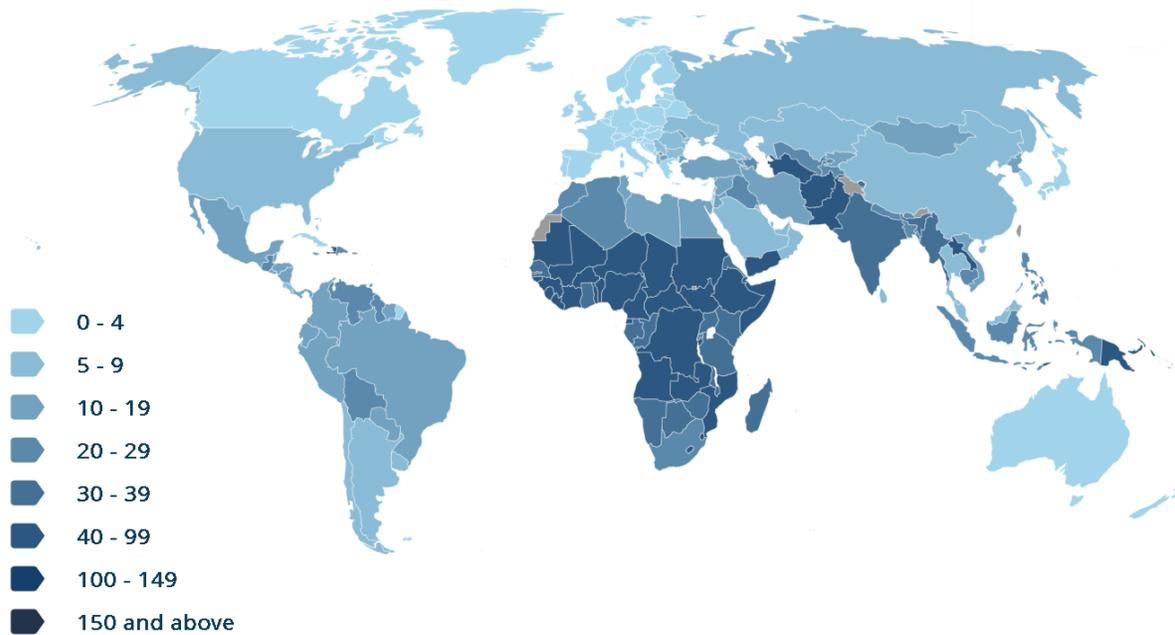
deterioration in the left tail. In fact, the left tail improves as well. More broadly, the whole distribution seems to shift to the right (compared to the control group). In Panel K, we observe a similar pattern when looking at the post-treatment distributions of the share of live births. (A consistent 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).

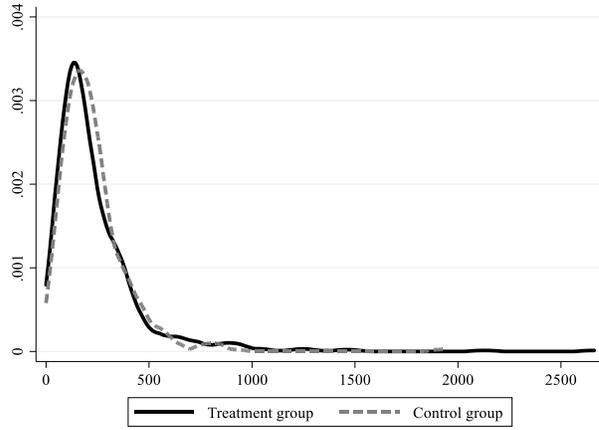
Figure A1. Infant mortality rate



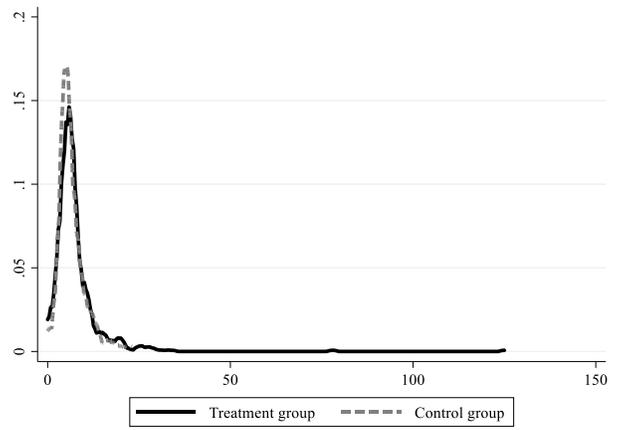
Notes. Infant mortality rate is measured by the number of deaths of infants under one-year old per 1,000 births alive.
Source: United Nations Inter-Agency Group for Child Mortality Estimation (<https://childmortality.org/data>).

Figure A2. Pre-treatment distributions

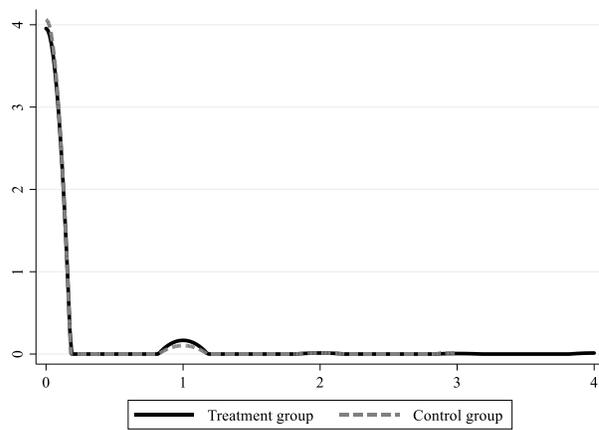
A. Primary healthcare services per employee at $t - 1$



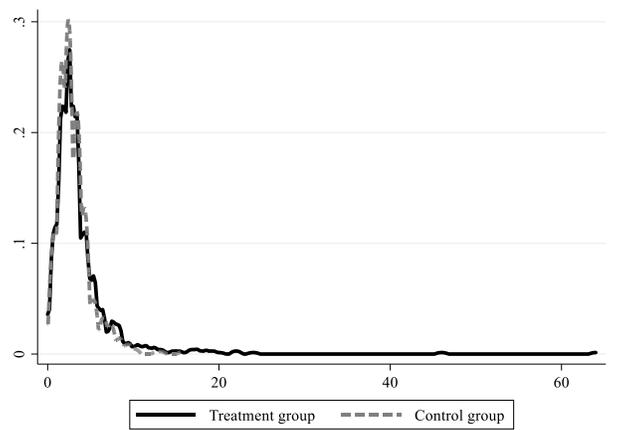
B. Employee at $t - 1$



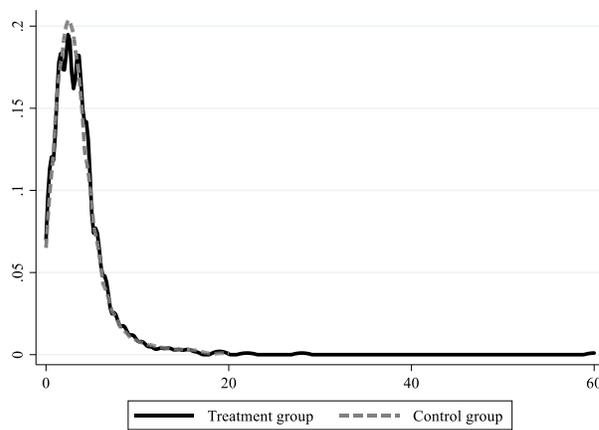
C. Doctors at $t - 1$



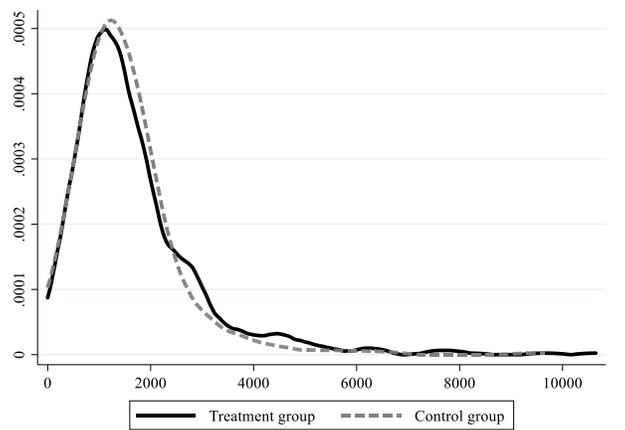
D. Nurses at $t - 1$



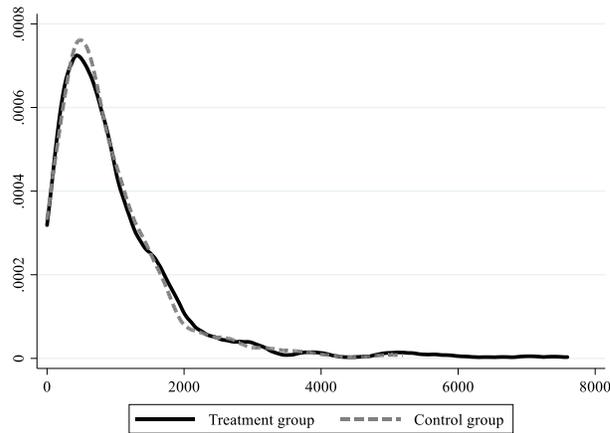
E. Administrative employees at $t - 1$



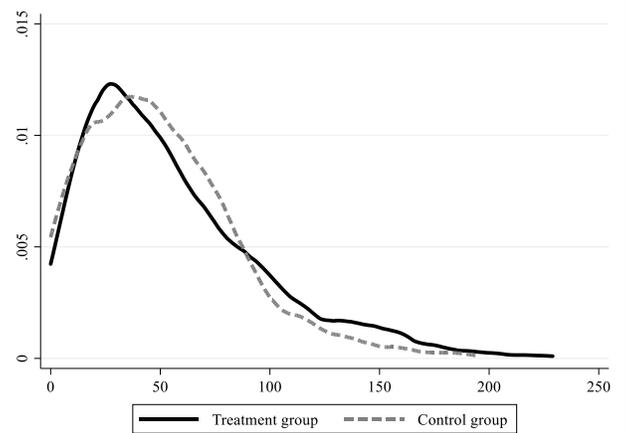
F. Primary healthcare services at $t - 1$



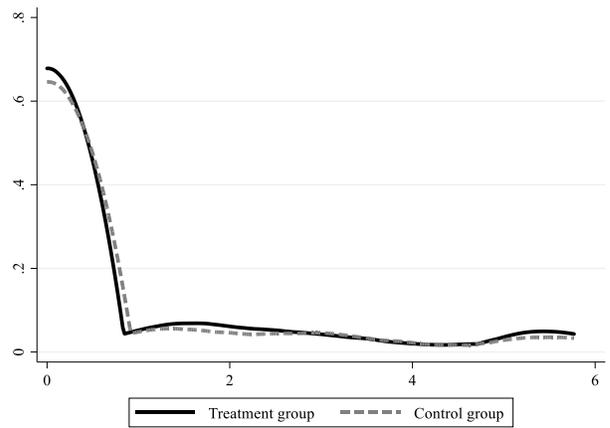
G. Maternal and childhood healthcare services at $t - 1$



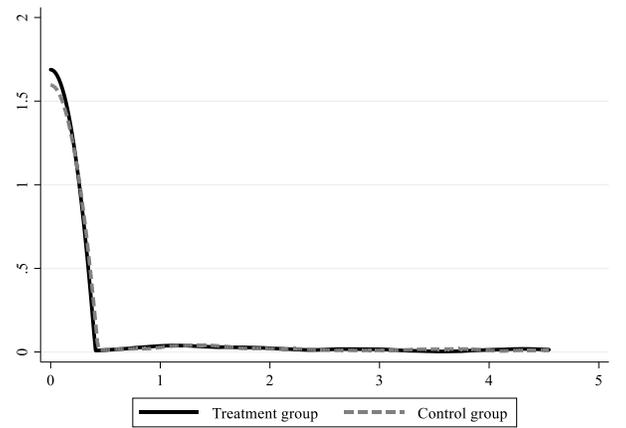
H. Births at $t - 1$



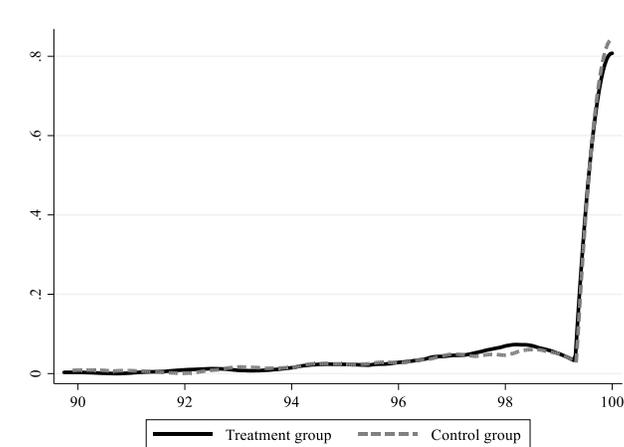
I. Share of stillbirths at $t - 1$



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

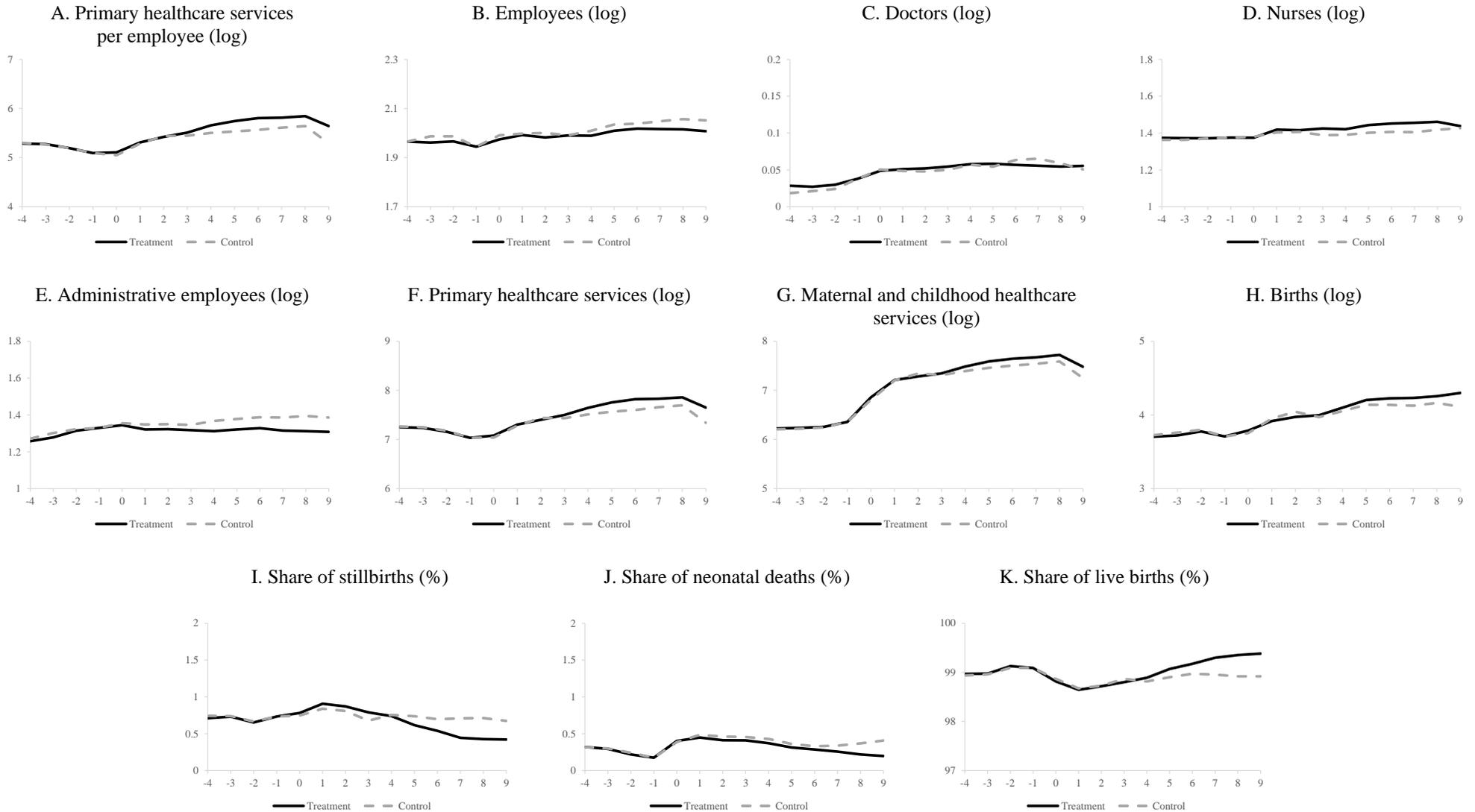


K. Share of live births at $t - 1$



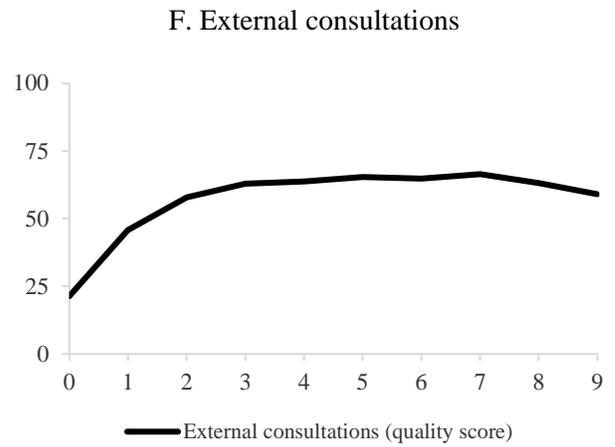
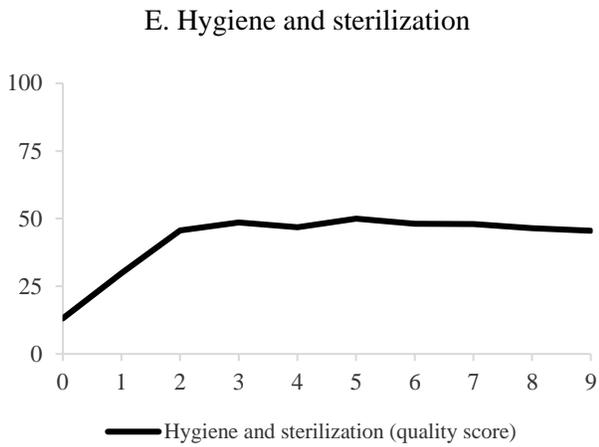
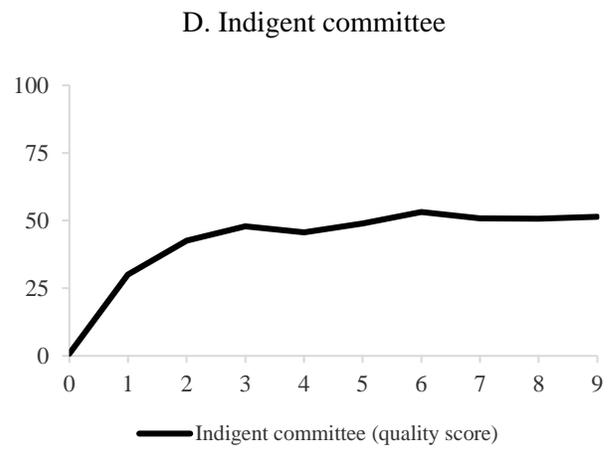
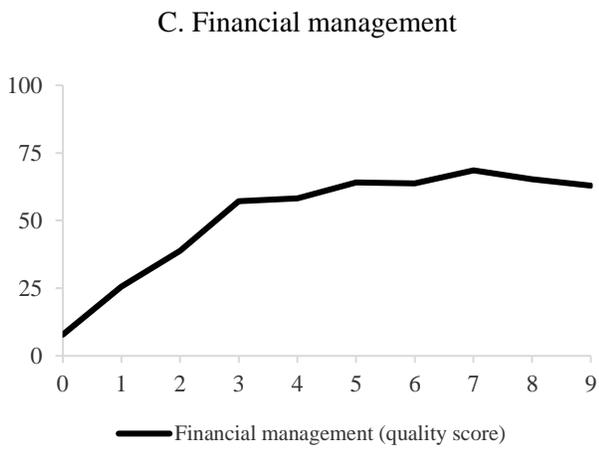
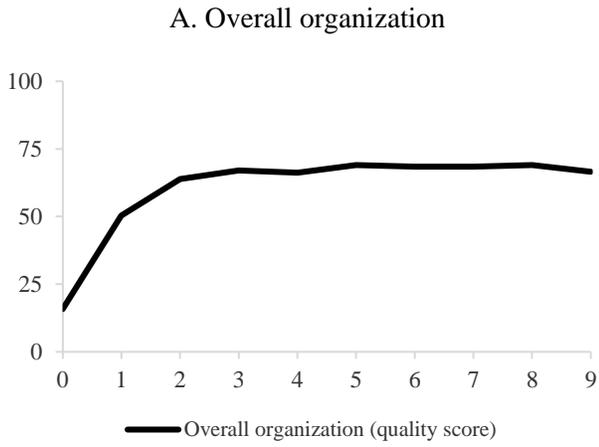
Notes. This figure plots the kernel density of each outcome variable at $t - 1$ (Q4 2016), that is, in the quarter that precedes the PDSS intervention, separately for the treatment (black solid line) and control (gray dashed line) groups.

Figure A3. Dynamics

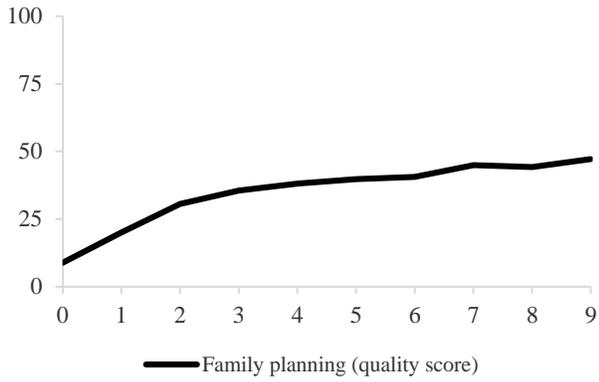


Notes. This figure plots the evolution of each outcome variable in the treatment and control groups. In each panel, the y-axis refers to the mean of the respective outcome variable across all health centers in the treatment (black solid line) and control (gray dashed line) groups, respectively. The x-axis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

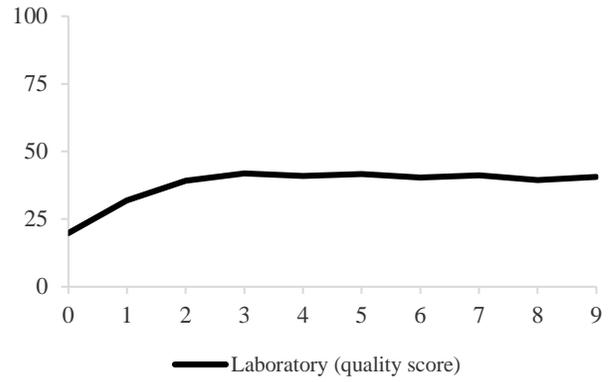
Figure A4. Quality improvements at the treated health centers



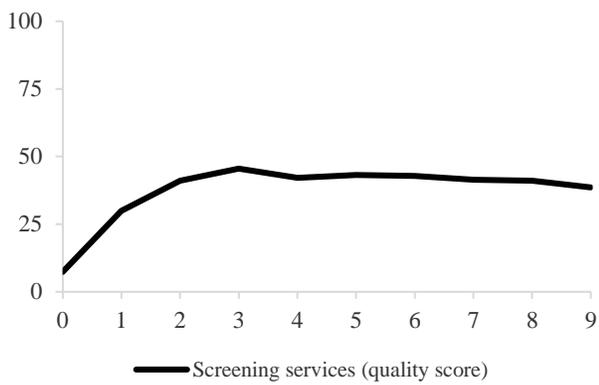
G. Family planning



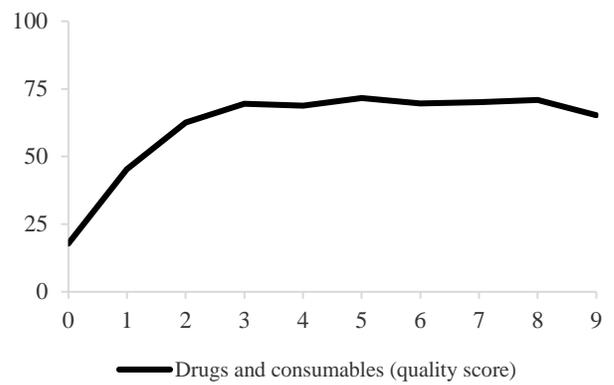
H. Laboratory



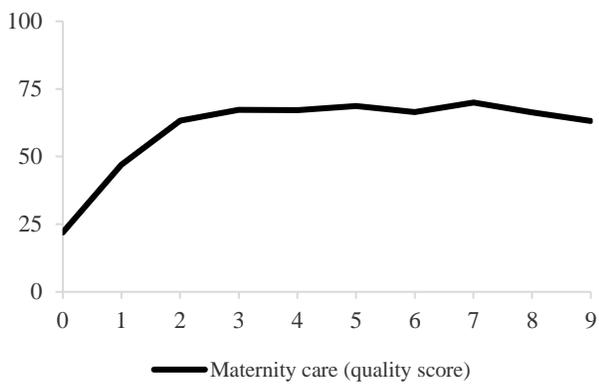
I. Screening services



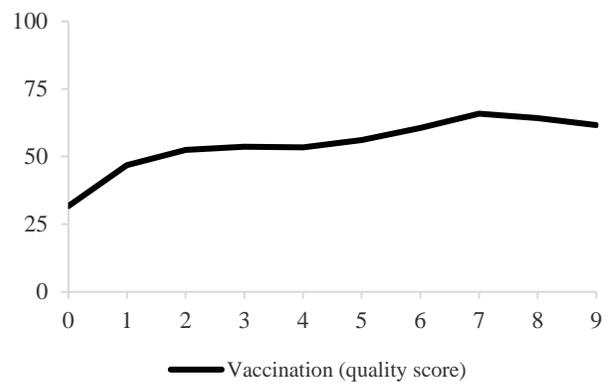
J. Drugs and consumables



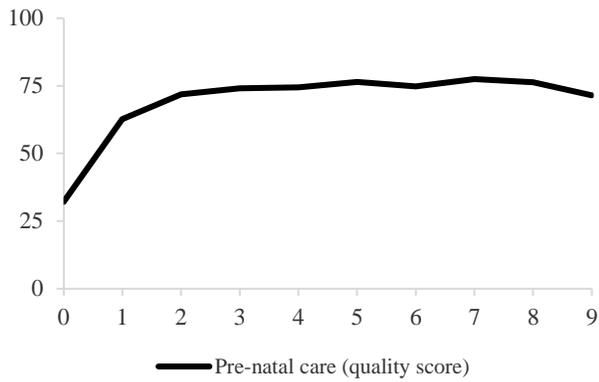
K. Maternity care



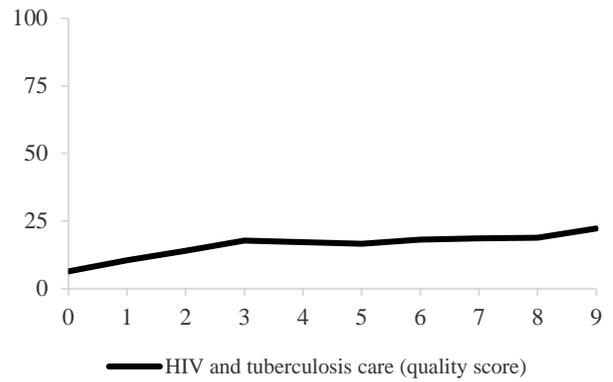
L. Vaccination



M. Pre-natal care



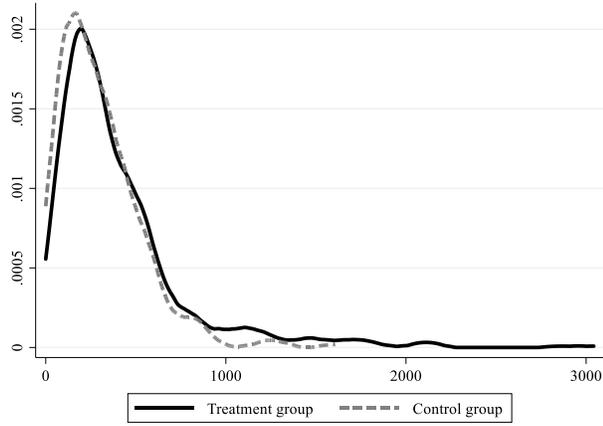
N. HIV and tuberculosis care



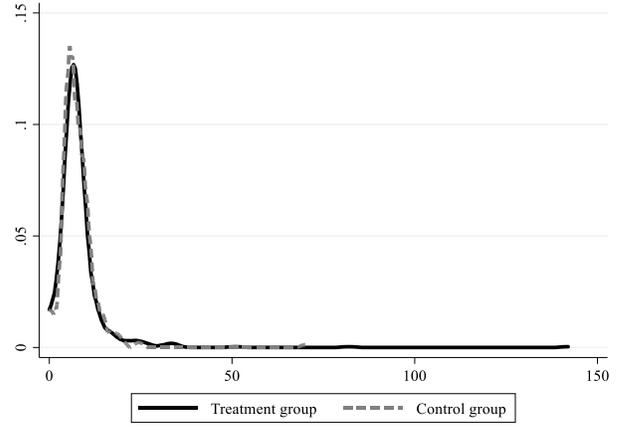
Notes. This figure plots the mean of the treated health centers' quality scores on a quarterly basis from $t = 0$ (Q1 2017) until $t = 9$ (Q2 2019). In each panel, the y -axis refers to the mean of the respective quality score. The x -axis indicates quarters in event time, where $t = 0$ is the initial 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

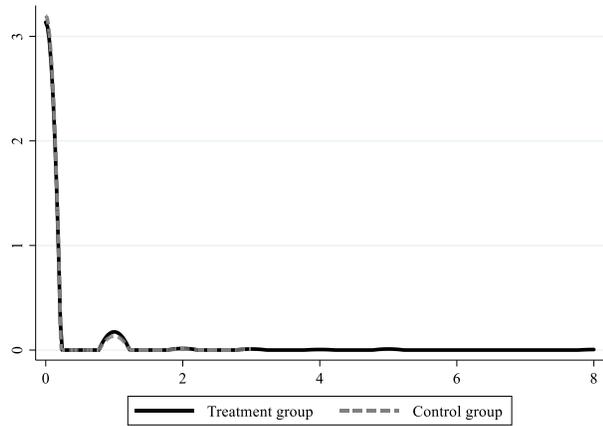
A. Primary healthcare services per employee at $t + 9$



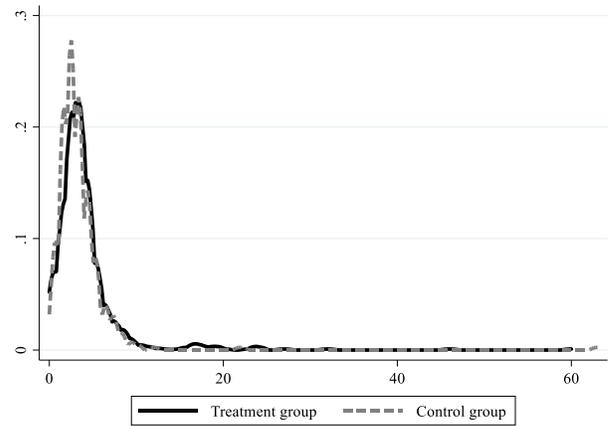
B. Employee at $t + 9$



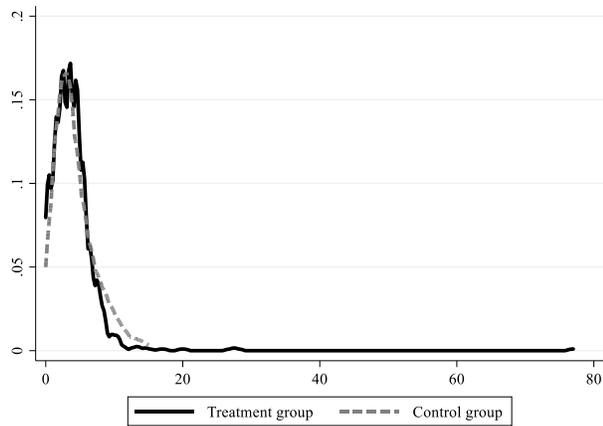
C. Doctors at $t + 9$



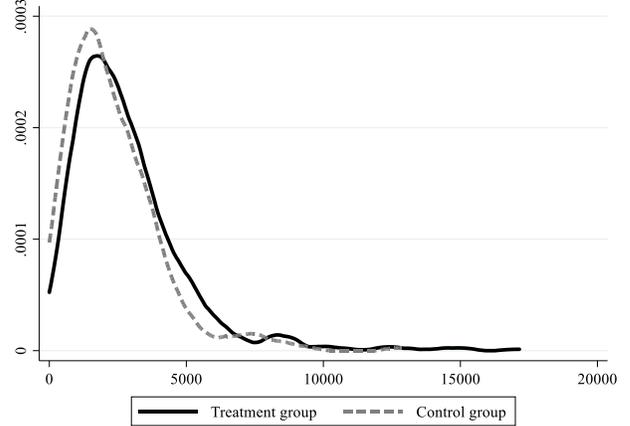
D. Nurses at $t + 9$



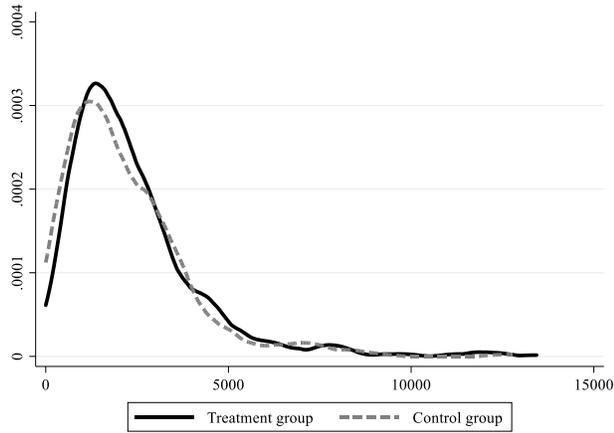
E. Administrative employees at $t + 9$



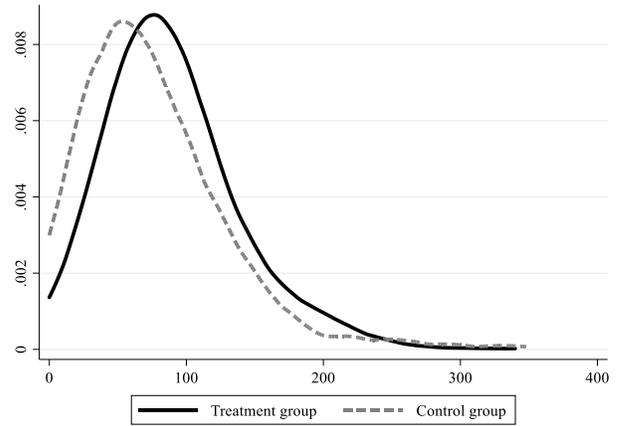
F. Primary healthcare services at $t + 9$



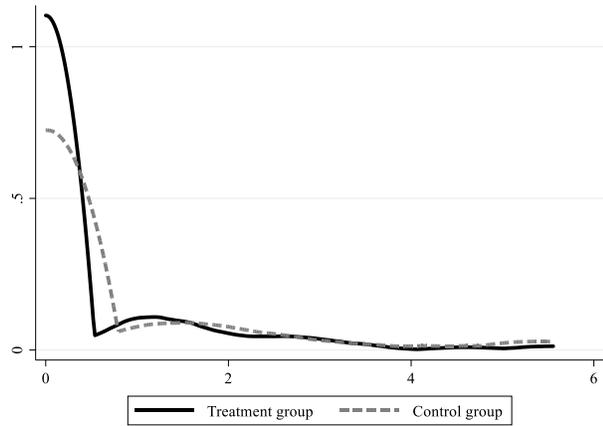
G. Maternal and childhood healthcare services at $t + 9$



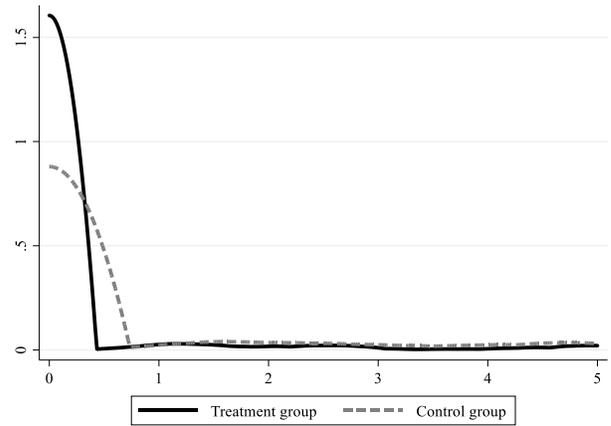
H. Births at $t + 9$



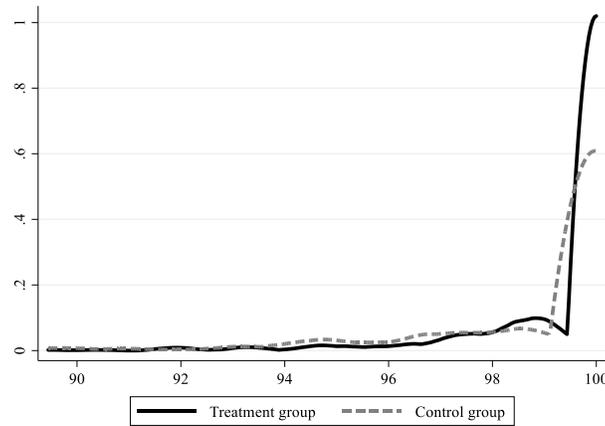
I. Share of stillbirths at $t + 9$



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+ pregnant 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 \times 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 (0.042)	-0.003 (0.017)	-0.004 (0.008)	0.012 (0.028)	-0.021 (0.024)	0.065 (0.045)	0.017 (0.062)	0.023 (0.041)	0.065 (0.070)	-0.031 (0.050)	-0.034 (0.095)
Treatment × post ($t = 5$ to 9)	0.255*** (0.060)	-0.022 (0.021)	-0.008 (0.009)	0.032 (0.036)	-0.058** (0.028)	0.233*** (0.068)	0.149** (0.075)	0.138*** (0.050)	-0.209*** (0.074)	-0.107** (0.051)	0.316*** (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 \times post ($t = 0$ to 4)	0.068 [0.201]	-0.003 [0.879]	-0.004 [0.626]	0.012 [0.705]	-0.021 [0.453]	0.065 [0.248]	0.017 [0.817]	0.023 [0.626]	0.065 [0.453]	-0.031 [0.626]	-0.034 [0.770]
Treatment \times post ($t = 5$ to 9)	0.255*** [0.001]	-0.022 [0.453]	-0.008 [0.453]	0.032 [0.453]	-0.058 [0.120]	0.233*** [0.007]	0.149 [0.126]	0.138*** [0.016]	-0.209** [0.016]	-0.107 [0.120]	0.316*** [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 (%) (1)	Share of neonatal deaths (%) (2)	Share of live births (%) (3)
Treatment × post ($t = 0$ to 4)	0.060 (0.070)	-0.026 (0.055)	-0.035 (0.095)
Treatment × post ($t = 5$ to 9)	-0.247*** (0.074)	-0.155*** (0.060)	0.402*** (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 \times post ($t = 0$ to 4)	0.068 (0.042)	0.065 (0.069)	-0.031 (0.052)	-0.034 (0.095)
Treatment \times post ($t = 5$ to 9)	0.242*** (0.052)	-0.200*** (0.066)	-0.105** (0.052)	0.305*** (0.092)
Treatment \times post ($t = 5$ to 9) \times Density of health centers	-0.105*** (0.034)	0.071** (0.033)	0.055 (0.038)	-0.126** (0.059)
Treatment \times post ($t = 5$ to 9) \times Infant mortality	-0.027 (0.025)	-0.493*** (0.039)	-0.118*** (0.040)	0.611*** (0.066)
Treatment \times post ($t = 5$ to 9) \times Size	-0.048 (0.033)	0.044 (0.026)	-0.011 (0.024)	-0.032 (0.041)
Treatment \times post ($t = 5$ to 9) \times Operating efficiency	-0.497*** (0.045)	-0.003 (0.023)	-0.038* (0.021)	0.041 (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 \times 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) (1)	Primary healthcare services per employee (log) (2)	Share of live births (%) (3)	Share of live births (%) (4)
Treatment × post ($t = 0$ to 4)	0.068 (0.042)	0.068 (0.042)	-0.034 (0.095)	-0.034 (0.095)
Treatment × post ($t = 5$ to 9)	0.113** (0.052)	0.159*** (0.052)	0.181* (0.096)	0.237** (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 (%) (1)	Share of neonatal deaths (%) (2)	Share of live births (%) (3)	Share of stillbirths (%) (4)	Share of neonatal deaths (%) (5)	Share of live births (%) (6)	Share of stillbirths (%) (7)	Share of neonatal deaths (%) (8)	Share of live births (%) (9)
Post ($t = 0$ to 4)	0.020 (0.184)	0.007 (0.094)	-0.027 (0.219)	-0.097 (0.177)	0.123 (0.124)	-0.026 (0.246)	0.024 (0.024)	0.006 (0.020)	-0.030 (0.034)
Post ($t = 5$ to 9)	0.032 (0.281)	0.036 (0.145)	-0.067 (0.320)	-0.064 (0.198)	0.110 (0.172)	-0.046 (0.293)	0.033 (0.041)	-0.003 (0.033)	-0.030 (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.	<i>p</i> -value (diff. in means)	Obs.	<i>p</i> -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.