ERF 2020

26TH Annual Conference

Accreditation as a Quality-Improving Policy Tool: Family Planning, Maternal Health, and Child Health in Egypt

Amira El-Shal, Patricia Cubi-Molla and Mireia Jofre-Bonet

ECONOMIC RESEARCH FORUM
منتدى البحوث الاقتصادية

Accreditation as a Quality-Improving Policy Tool: Family Planning, Maternal Health, and Child Health in Egypt

Amira El-Shal [a,b,1], Patricia Cubi-Molla [c], Mireia Jofre-Bonet [c,d]

[a] African Development Bank, Côte d'Ivoire
[b] Cairo University, Egypt
[c] Office of Health Economics, United Kingdom
[d] City, University of London, United Kingdom

Abstract

Accreditation of healthcare providers has been established in many high-income countries and some low- and middle-income countries as a tool to improve the quality of health care. However, the available evidence on the effectiveness of this approach is limited and of questionable quality, especially in low- and middle-income countries. In light of the interventions introduced under Egypt's Health Sector Reform Program between 2000 and 2014, we estimate the effect of health facility accreditation on family planning, maternal health, and child health outcomes using difference-in-differences fixed-effects and propensity score matching difference-in-differences models. To do so, we *spatially* link women to their nearest mapped health facilities using their Global Positioning System coordinates. We find that accreditation had multiple positive effects, especially on delivery care and child morbidity prevalence. The positive effects of accreditation appear to weaken over time though. Our findings suggest that facility accreditation can be effective in improving family planning, antenatal care, delivery care, and child health, but demand the study into how the positive effects can be sustained.

Keywords: Health, health behavior, healthcare, managed care, morbidity, outpatient, public health, regulation.

JEL classification: I11; I12; I18.

[1] Corresponding author: a.elshal@afdb.org,
African Development Bank, Avenue Joseph Anoma, 01 BP 1387 Abidjan 01, Côte d'Ivoire.

1. INTRODUCTION

In 1997, Egypt launched a comprehensive Health Sector Reform Program (HSRP) to address fundamental challenges in the healthcare system. The HSRP introduced an integrated package of service delivery and financing interventions to address the means by which primary health care (PHC) is financed, delivered, organized, and managed. One of the cornerstones of the HSRP was the facility accreditation program, which was defined as a process for evaluating PHC facilities according to a set of standards that define activities and structures that directly contribute to improved patient outcomes. The main aim of the program was to provide the HSRP with a framework for continuous quality improvement.

Accreditation of healthcare providers has been established in many high-income countries, and some low- and middle-income countries, as an approach to improve the quality of care that combines the two elements of quality assurance and quality improvement (Hort *et al.*, 2013). There is consistent evidence that shows that accreditation programs improve the process of care provided by healthcare services (Alkhenizan & Shaw, 2011). However, there is limited evidence on the effectiveness of accreditation on patient-level outcomes in all settings (Brubakk *et al.*, 2015; Flodgren *et al.*, 2016; Hinchcliff *et al.*, 2012). Moreover, in light of evidence on the non-monotonic effect of access to health care on gender inequality (Oster, 2009), it is particularly important to estimate the effect of having access to an accredited facility on disadvantaged groups, such as women and children. To date, no study in a low- or middle-income country investigated the effect of quality improvement through accreditation on key patient outcomes. Since accreditation usually entails significant costs, investigating its effectiveness is crucial, especially in settings where resources are extremely constrained.

We attempt to fill the gap in the literature by investigating the effect of quality improvement through accreditation on patient outcomes in Egypt, a middle-income country. The paper exploits the quasi-natural experiment associated with the introduction of Egypt's facility accreditation program to estimate the effect of having access to an accredited facility on a set of family planning, maternal health, and child health outcomes. To do this, difference-in-differences (DiD) fixed-effects models are used for the period 1992-2014. DiD is also combined with propensity score matching (PSM) for the specific periods of 2000-2005, 2005-2008, and 2008-2014.

The paper is organized as follows: section 2 provides some background; section 3 explains our empirical strategy; section 4 describes the data we exploit; section 5 presents and discusses the results, and section 6 concludes. The Appendix contains descriptive statistics, robustness checks, and extensions.

2. BACKGROUND

2.1 The facility accreditation program

In 1997, the Government of Egypt (GOE) launched the HSRP, a new PHC strategy to reform the health system in phases over a period of 15-20 years. The program came into operation in 2000. The simultaneous implementation of the program across the country was deemed infeasible due to pre-existent constraints in the healthcare system and the complex nature of interventions to be introduced. Hence, GOE decided to implement the program over phases. The targeting took place at the district level in the participating governorates, whose master plans relied on a social vulnerability index to target districts of the most vulnerable populations. Early entrants to the program included a group of PHC facilities in the governorates of Alexandria, Menoufia, and Sohag, which represent, respectively, urban governorates, Lower Egypt, and Upper Egypt. The three pilot governorates represent the three major regions in Egypt. Other governorates followed subsequently.

One of the key marketing points of the PHC model of service provision introduced under the HSRP is that the model would improve access to quality care. To ensure this, the facility accreditation program was introduced by the Ministry of Health (MOH) as a process to monitor and facilitate the quality of services and influence the behavior of healthcare providers. Thus, as part of the HSRP, accreditation became obligatory to all PHC facilities in the districts targeted by GOE to join the program. Technical assistance was provided to develop improvement plans.

In the preparatory phase of the facility accreditation program, a PHC facility must meet the following criteria: (i) has a process to monitor, evaluate, and improve the quality of care; (ii) has a patient record system; (iii) provides a defined package of services including reproductive health obstetrics and gynecology, neonatal care, pediatric and adult medical care, basic emergency care, preventive health services, and ambulatory care; (iv) has been in operation for at least six months, has appropriate license by MOH and relevant medical union, and operates within the laws and regulations.

The survey is a key step in the facility accreditation program. A site visit to PHC facilities is conducted by a team of experts trained in accreditation using pre-set accreditation survey instruments and tools. The purpose of the accreditation survey is to evaluate the extent to which a facility complies with the nationally established accreditation standards, and accordingly, determine whether it is awarded or denied accreditation. The assessment includes eight categories: patient rights, patient care, safety, management of support services, management of information, quality improvement program, family practice, and management of the facility. Optimal standards in each category focus on key processes, activities, or outcomes that facilities should achieve.

During the accreditation survey, trained surveyors use three approaches to collect data and measure compliance with the established standards: review of specific administrative and clinical records; observe the performance of specified tasks in particular areas; and conduct personal interviews. If a facility scores 80 percent or above in the total survey scores, it is granted full accreditation for a period of two years. If a facility scores between 50 and 79 percent of the total survey score, it is granted provisional accreditation for one year, after which a reassessment survey is conducted to investigate if the deficits pointed out by the first round have been addressed. If a facility scores less than 50 percent, accreditation is denied. The scoring by areas is presented in Tables A.1 and A.2 in Appendix A.

In terms of contribution to the total score, the quality dimension *patient care* contributes the most. This dimension measures the extent to which patients receive appropriate care, and focuses on compliance with clinical practice guidelines as well as appropriate diagnosis. Details of the patient care assessment process are included in the Insert A.3 in Appendix A.

2.2 Anticipated effect of the facility accreditation program

One important characteristic of healthcare markets is the presence of asymmetric information (Arrow, 1963). It is well known that healthcare providers may act as “imperfect” agents of patients and over or under provide care or fail to deliver the adequate health care quality, which can become a health concern. Different interventions, including accreditation of providers, have evolved in response to these problems. By subjecting healthcare providers to a formal process that makes them meet pre-determined standards, accreditation is expected to minimize variations in medical practice, eliminate medically inappropriate care, control costs, and address the possibility that quality is underprovided (Akerlof, 1970; Viswanathan & Salmon, 2000).

In the context of this study, accreditation is expected to have a direct effect on some maternal health, child health, and family planning outcomes, through improving the quantity and quality of pertinent health services provided, and an indirect effect on other outcomes. Accreditation of health facilities certifies high compliance with standards defining activities and structures that directly contribute to improved patient outcomes. Hence, within the quality dimension *patient car*e, accreditation standards established to measure compliance of facilities in the subareas of antenatal care (ANC), Integrated Management of Child Illnesses (IMCI), and family planning are expected to have a primary effect on ANC coverage (number of visits), quality of ANC (being informed of signs of pregnancy complications, weight measurement, blood pressure measurement, and urine sample collection), child morbidity prevalence (acute respiratory infection (ARI), fever, and diarrhea), and informed choice of contraceptive methods (knowledge of side effects of contraceptive method used and knowledge of other methods of contraception that could be used). These outcomes reflect some of the standards assessed during the accreditation survey (see Appendix A, Insert A.3). We expect improvements in these outcomes in accredited compared to non-accredited facilities.

In parallel, accreditation is expected to have a secondary effect on the utilization of antenatal and delivery care services. Quality improvement in accredited facilities introduces an incentive for individuals to seek care at these facilities. The effect of this incentive is expected to be more significant in the sub-areas of care included in the assessment of the accreditation survey. Thus, we expect having access to an accredited facility to be associated with higher ANC coverage, higher institutional birth-delivery, and higher skilled assistance during delivery. This expectation holds given that accredited facilities were not functioning at full capacity prior to accreditation and can increase supply in the short term.

2.3 Evidence on the effect of accreditation

There exists a large body of literature on the effects of accreditation as a quality signaling device for firms (for instance, a good summary on firm behavior and accreditation can be found in Dranove & Jin, 2010). However, less is known on the impact of accreditation on health care provision. The majority of studies on accreditation in health care report on its positive effects on compliance with quality standards (Al Tehewy *et al.*, 2009; Bukonda *et al.*, 2002; Hong *et al.*, 2011; Salmon *et al.*, 2003). As per accreditation in low- and middle-income countries, a quasi-experimental study in Egypt found that accredited non-governmental health units had higher compliance with quality standards compared with non-accredited units (Al Tehewy *et al.*, 2009). Another study in Egypt on public clinics found that providers that had earned an accreditive *Gold Star* were more likely to adhere to higher quality practices in counseling and examination than non-Gold Star facilities (Hong *et al.*, 2011). According to a descriptive study in Zambia, a national hospital accreditation program was associated with significant improvement in compliance of accredited hospitals with standards in overall scores and in seven out of 13 functional areas (Bukonda *et al.*, 2002). In South Africa, Salmon *et al.* (2003) used a randomized control trial to investigate the effect of an accreditation program on public hospitals and found that the processes' and outcomes' average compliance of accredited hospitals improved significantly, while no significant increase was observed in non-accredited hospitals.

Besides compliance with standards, the majority of the studies report on the effect of accreditation on quality of care measures. These are, for the most part, not patient health outcomes, but downstream process indicators (Al Tehewy *et al.*, 2009; El-Jardali *et al.*, 2008; Hong *et al.*, 2011; Quimbo *et al.*, 2008; Salmon *et al.*, 2003). Unlike compliance with standards, there is no conclusive evidence on the effect of accreditation on quality of care. While El-Jardali *et al.* (2008), Hong *et al.* (2011), and Quimbo *et al.* (2008) report a positive effect of accreditation on different indicators of quality of care, the studies employing more robust study designs report mixed effects. These are Salmon *et al.* (2003) and Al Tehewy *et al.* (2009), which used a randomized controlled trial and a quasi-experimental design, respectively. In a study based on data from hospitals in South Africa, Salmon *et al.* (2003) found limited or no effect of a randomized accreditation program on quality measures apart from increases in perception of quality among nurses. In Egypt, Al Tehewy *et al.* (2009) found a positive effect of accreditation of non-governmental health units on patient satisfaction with respect to all areas of health service (cleanliness, waiting area, waiting time, and staff performance). As for provider satisfaction, the study found a positive effect on the overall satisfaction score, but no significant difference in the mean satisfaction score between the accredited and non-accredited units with respect to the social environment, administrative environment, and family health model.

In conclusion, the available evidence on the effectiveness of quality improvement through accreditation of healthcare providers is limited and of questionable quality but suggests that accreditation could improve the process of care provided by different providers. However, evidence is limited on the effectiveness of accreditation on patient outcomes. We could not identify any study of the effect of accreditation on key patient outcomes such as family planning, maternal health and child health outcomes.

3. METHODS

3.1 Difference-in-differences

As the accreditation policy is staggered, we first follow Wooldridge (2012) and use a general DiD fixed-effects model to estimate the effect of accreditation by comparing the health outcomes of accredited facilities (treatment group) to that of non-accredited facilities (control group) before and after accreditation (treatment) between 1992 and 2014. Accredited and non-accredited facilities presumably differ in observed characteristics, such as labor force, and unobserved ones too, such as managerial ability. The DiD method controls for both observed as well as unobserved characteristics that are time invariant. Fixed effects further eliminate any confounding that might be caused by facility effects, whether observed or unobserved, which are constant over time within each facility. With regard to time-varying unobservable factors, we report in Appendix C the results of the parallel-trends test, which provide evidence of the absence of unobserved time-varying confounding, establishing the unbiasedness of our DiD estimates.

Treatment self-selection is not a concern in our context as treatment, i.e., the decision of whether or not to obtain accreditation is exogenous. Accreditation is not a function of some choice of the treated unit, but is rather a function of a policy that differentially affects units based on pre-determined characteristics. As noted earlier, the facility accreditation program is only rolled out in certain geographical areas. Hence, whether one facility can obtain accreditation or not is neither self-selected nor screened.

For each health facility i at time t, we estimate the following DiD fixed-effects model:

$$y_{it} = \alpha + \beta\, policy_{it} + \gamma\, year_t + \delta_i + \varepsilon_{it} \qquad (1)$$

where y_{it} denotes a health outcome of interest y for facility i at time t; $policy_{it}$ is the key policy variable that equals one if facility i is subject to accreditation in year t; $year_t$ is a time-period dummy; and δ_i is the unobserved facility effect.

Additionally, to compare between the effects of accreditation for the specific periods of 2000-2005, 2005-2008, and 2008-2014, we estimate the typical DiD specification below for each health facility i at time t.

$$y_{it} = \alpha + \beta\, acc_{it} + \gamma\, d_{post} + \delta\, acc_{it} * d_{post} + \zeta\, fac_i + \eta\, dist_i + \varepsilon_{it} \qquad (2)$$

where t equals 0 for the baseline years (2000, 2005, or 2008) and 1 for the follow-up years (2005, 2008, or 2014); acc_{it} is an indicator variable that takes value 1 if facility i is accredited and 0 if not; d_{post} is an indicator variable for the follow-up year; the interaction term $acc_{it} * d_{post}$ measures the effect of accreditation in the follow-up year; and δ, our main coefficient of interest, captures the effect of accreditation on the outcome at the facility level. The vector fac_i contains facility-level controls that reflect different characteristics of facility i; and $dist_i$ is a vector of district-level controls including social, economic, and demographic characteristics of the district in which facility i is located.

For each health outcome, y_{it}, we report the results of three study periods: 2000-2005, 2005-2008, and 2008-2014.

3.2 Propensity score matching difference-in-differences

The targeting of the reform interventions at the district level under the HSRP followed a socio-economic vulnerability index of the areas around them. As such, the comparison of health outcomes without accounting for this would be biased. To ensure that no bias exists due to targeting, we combine DiD with the PSM approach.[2] Matching on *observables* mitigates the potential bias by pairing accredited and non-accredited health facilities based on pre-accreditation *observable* characteristics, which were initially used by GOE for accreditation targeting. Additionally, as a stand-alone method, DiD can be used to identify treatment effects if there is a selection based on (time-invariant) *non-observables*. Thus, while conventional PSM cannot account for *non-observables*, combining matching with DiD enables us to account for both the effect of *observable* and *unobservable* heterogeneity that is constant over time as well as the targeting policy. To minimize any potential bias due to time-varying unobservable factors, we also control for an extensive set of facility-level characteristics as well as population coverage of the facilities. Reassuringly, we generally find no significant differences in outcomes or characteristics between the population covered by treated and non-treated facilities (see Appendix C, Table C.4). Table C.4 suggests that our matching strongly satisfies the requirement of conditional independence.

To obtain the PSM DiD estimates, we follow Rosenbaum & Rubin (1983). We first apply PSM to match facilities and then extend the conventional DiD estimate by defining health outcomes conditional on propensity scores and applying semi-parametric methods to construct the differences. First, we match treated and control health facilities based on pre-treatment observable characteristics and use Kernel functions to assign weight to the jth control

---

[2] The mixture of PSM DiD was first proposed by Heckman *et al.* (1998). PSM DiD estimates are superior to the conventional DiD estimates as no functional form restrictions are imposed when estimating the conditional expectation of the outcome variable using PSM DiD.

facility matched to the *i*th treated facility.[3] As such, in our context, the propensity score is the probability of being targeted by the HSRP intervention given a set of observable social and economic indicators used to construct the socio-economic vulnerability index. Second, we estimate a DiD specification in equation (2) with health outcomes defined conditional on the propensity score generated earlier. The Kernel PSM DiD estimate for each treated facility i is calculated as

$$\hat{\delta}_i = \left(y^T_{ipost} - y^T_{ipre}\right) - \sum_{j\in C} \omega(i,j)\left(y^C_{jpost} - y^C_{jpre}\right) \qquad (3).$$

Prior to the DiD estimation, we verify that the *common support assumption* is satisfied by checking the overlap between treatment (accredited facilities) and control (non-accredited facilities) groups (see Appendix C, Table C.4). Once the matching is applied, we use two-sample t-tests to examine if there are significant differences in the means of observable characteristics for both groups (Rosenbaum & Rubin, 1983).

While we use district-level social and economic indicators to estimate the propensity score, facility-level characteristics are used as additional covariates later in the DiD estimations. For each of our health outcomes, we report the results for three study periods: 2000-2005, 2005-2008, and 2008-2014.

4. DATA

4.1 Dependent variables

We make use of all the relevant data made available by the Egypt Demographic and Health Survey (DHS) on family planning and maternal and child health. This is the data that we expect to reflect the effect of compliance with quality standards, policies and procedures, which are the focus of accreditation assessment. Our dependent variables are outcomes of informed

---

[3] Using weights: $\omega(i,j) = \frac{K(\frac{P_j - P_i}{a_n})}{\sum_{K\in C} K(\frac{P_k - P_i}{a_n})}$, where a is the bandwidth parameter; K is the Kernel function and P_i and P_j are the propensity scores for treated and control facilities.

choice of contraceptive methods, ANC, delivery care, and child morbidity prevalence.

We collapse the individual responses of 97,990 women over the period 1992-2014 at the facility level, drawing from six DHS waves: 1992, 1995, 2000, 2005, 2008, and 2014. To do so, we use the GPS coordinates of both women interviewed and health facilities to link each woman to the nearest mapped facility for each wave of the Egypt DHS. Thus, we identify women who live in the catchment area of accredited facilities (treatment group) and those in the catchment area of non-accredited facilities (control group). Then, we calculate health outcomes at the facility level for each of the Egypt DHS waves and combine outcomes in a panel. All eligible PHC facilities across Egypt are used during the joining process.

Family planning. As part of the family planning sub-area of the patient quality of care dimension, the accreditation surveyor checks if the facility has a good information/education/communication (IEC) system. For this subarea, we include two family planning outcomes that capture the effect of accreditation on informed choice of contraceptive methods by calculating the percentage of current users of selected contraceptive methods who were informed of the side effect or problems of the method used.[4] We also calculate the percentage of current users of selected contraceptive methods who were informed of other methods of contraception that could be used. Informed choice emphasizes that women choose the method that best satisfies their personal and reproductive health needs based on a thorough understanding of other methods of contraception they could use.

ANC. As part of the ANC sub-area of the patient quality of care dimension, the surveyor checks if physical examination is performed for all patients. We include six ANC outcomes that capture the effect of accreditation on the quality of ANC. We calculate the percentages of mothers who received the

[4] Those receiving information on the efficacy and side effect of contraceptives used tend to have higher continuation rates than those who do not (NCC, 2005).

following components of ANC: being informed of signs of pregnancy complications, weight measurement, blood pressure measurement, and urine sample collection. As part of the accreditation survey, the surveyor checks as well if the number of ANC visits falls within the clinical guidelines. Therefore, we calculate an outcome of ANC coverage indicator (at least four visits). This outcome is used as a global preferred outcome of access to and use of health care during pregnancy to track performance in maternal health programs.[5] A pregnant woman is expected to receive health interventions during antenatal visits that could be vital to her health and the health of her infant as well.

Delivery care. As part of the ANC sub-area of the patient quality of care dimension assessed by the accreditation survey, the surveyor assesses patient's knowledge and understanding of delivery services provided in the facility. The Egypt DHS allows us to calculate two delivery care outcomes to capture the effect of compliance with the accreditation standards in this regard: institutional delivery and skilled assistance during delivery. The two outcomes are widely advocated for reducing maternal, perinatal, and neonatal mortality. Institutional delivery captures the effect of accreditation on expanding access to childbirth facilities and, more importantly, is a proxy measure of maternal and neonatal morbidity and mortality.[6]

The second but most important measure of delivery care included in our analyses is skilled assistance during delivery. Empirical literature provides evidence that wider access to professional care during pregnancy and childbirth reduces maternal mortality. Women assisted by skilled health personnel during delivery are less likely to die from any cause related to or aggravated by childbirth (Graham *et al.*, 2001).

[5] WHO recommends that a woman receives at least four antenatal visits during a normal pregnancy to ensure that antenatal complications are detected and controlled at the earliest stage.

[6] Women who give birth at a health facility are more likely to receive proper medical attention and care during delivery, as do their infants.

Child morbidity prevalence. As part of the "IMCI" sub-area of the patient care dimension of quality assessed by the accreditation survey, the surveyor checks if child is checked for cough, diarrhea, sore throat, ear infection, and fever. We use the prevalence of childhood ARI, fever, and diarrhea from the Egypt DHS data as three outcomes reflecting morbidity prevalence.

ARI is the leading infectious cause of death in children worldwide.[7] Diarrheal diseases are the second leading cause of death in children under age five (World Health Organization, 2016). The risk of under-five mortality could be diminished substantially through reducing the prevalence of ARI and diarrheal diseases and encouraging women to seek treatment for their children at a health facility or from a healthcare provider.[8]

4.2 Explanatory variables

The explanatory variables included in the analyses of this study are a treatment variable that reflects participation in the facility accreditation program, facility-level controls, district-level social and economic controls, and regional dummies to control for regional variation.

Treatment and control facilities. The gradual uptake of the facility accreditation program by health facilities provides a quasi-natural experiment. We draw on facility-level data from MOH to categorize facilities as treated (accredited, either fully or provisionally) and non-treated (non-accredited). To ensure *treatment* reflects only accreditation, we remove from the sample accredited health facilities subject to additional interventions under the HSRP such as performance-based financing (PBF) and introducing user fees.

Matching. In order to eliminate potential unobserved heterogeneity and account for possible differences between accredited and non-accredited facilities prior to accreditation, we include a set of facility and district

---

[7] Mortality due to ARI accounted for 16 percent globally of the total deaths among under-five children in 2015.

[8] We are not able to calculate indicators of child morbidity treatment as observations in our sample are not statistically sufficient.

characteristics in our analyses such as labor force, the facility's building condition, and population coverage. For labor force, we incorporate the number of eight types of workers in a health facility: practitioners, specialists, pharmacists, nurses, lab technicians, x-ray technicians, health observers, and social workers. For the building infrastructure, a dummy variable that describes the condition of a facility building as 'bad', 'average,' or 'good'. As for population, we control for the size of population in the facility catchment area.

At the district level, we use Egypt's 2006 Population and Housing Census to construct the eight district-level social and economic controls: illiteracy ratio, unemployment ratio, income dependency ratio, inaccessibility to electricity, inaccessibility to potable water, average family size, household (HH) crowding factor, and population size. In addition, regional dummies are defined for fully urban governorates, urban Lower Egypt, rural Lower Egypt, urban Upper Egypt, rural Upper Egypt, and frontier governorates. These district-level covariates control for both the selection criteria of the HSRP targeting and the demographic variation across districts. As discussed earlier, the regional targeting of the HSRP follows a socio-economic vulnerability index that is constructed from the eight social and economic indicators outlined earlier.

The descriptive statistics of district characteristics of facilities based on which targeting took place are reported and discussed in Appendix B.

5. RESULTS AND DISCUSSION

Table 1 reports the DiD fixed-effects estimates of the effects of accreditation during the period 1992-2014. We additionally report the DiD and Kernel PSM DiD results for the specific periods 2000-2005, 2005-2008, and 2008-2014 in Table 2.

Table 1: Difference-in-differences fixed-effects estimates of the effects of accreditation, 1992-2014

	Family planning		ANC					Delivery care		Child morbidity prevalence		
	Knowledge of side effects	Knowledge of contraceptives	4+ visits	Informed of complications	Weight measurement	Blood pressure measurement	Urine sample collection	Institutional delivery	Skilled-assisted delivery	ARI	Fever	Diarrhea
treat=1	-1.168	1.318	0.717	3.305	-1.445	-2.710	-1.350	-1.523	0.226	-1.740	1.658	-2.674**
	(2.357)	(2.461)	(1.823)	(2.729)	(1.780)	(1.788)	(2.501)	(1.781)	(1.647)	(1.506)	(1.820)	(1.197)
Years (Ref: 1992)												
1995	-4.360**		6.288***					7.678***	7.559***	14.132***	19.168***	3.028**
	(1.990)		(1.938)					(2.007)	(2.004)	(1.286)	(1.719)	(1.178)
2000	37.991***	40.478***	15.990***					21.227***	18.313***	0.398	-3.950**	-5.935***
	(2.044)	(1.855)	(1.931)					(1.904)	(1.885)	(1.066)	(1.542)	(1.090)
2005	38.417***	47.627***	36.737***	12.471***	25.150***	23.544***	26.875***	39.178***	33.367***	2.308**	-0.573	4.236***
	(1.960)	(1.670)	(1.786)	(1.870)	(1.577)	(1.519)	(1.800)	(1.874)	(1.843)	(1.122)	(1.580)	(1.148)
2008	43.714***	51.344***	40.673***	15.131***	23.499***	23.435***	22.960***	42.058***	34.188***	2.688**	-7.025***	-3.924***
	(1.973)	(1.841)	(1.846)	(2.039)	(1.609)	(1.550)	(1.969)	(1.910)	(1.906)	(1.251)	(1.653)	(1.086)
2014	37.326***	51.250***	53.917***	24.522***	26.489***	31.460***	31.044***	54.279***	42.006***	6.863***	1.376	-0.109
	(2.151)	(1.950)	(1.881)	(2.139)	(1.700)	(1.634)	(1.955)	(1.992)	(1.988)	(1.368)	(1.756)	(1.257)
Constant	7.753***	5.522***	26.436***	19.129***	62.212***	62.062***	46.657***	30.473***	46.392***	9.287***	21.932***	13.829***
	(1.413)	(1.071)	(1.428)	(1.226)	(1.126)	(1.100)	(1.292)	(1.523)	(1.532)	(0.860)	(1.241)	(0.853)
Obs.	3,526	3,444	3,808	2,935	2,937	2,937	2,937	3,810	3,810	3,807	3,807	3,807

Each *column* represents a separate regression. Clustered standard errors are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Table 2: Difference-in-differences and Kernel propensity score matching difference-in-differences estimates of the effects of accreditation; 2000-2005, 2005-2008, and 2008-2014

	Outcome	2000-2005		2005-2008		2008-2014	
		DiD	PSM DiD	DiD	PSM DiD	DiD	PSM DiD
Family planning	Knowledge of side effects	16.853	15.777***	9.915**	1.246	-2.207	-3.586
		(10.306)	(3.990)	(4.275)	(3.309)	(2.859)	(3.465)
	Knowledge of contraceptives	8.370	6.578*	-0.402	-8.643***	1.622	0.356
		(8.253)	(3.528)	(4.014)	(3.222)	(3.080)	(3.560)
ANC	4+ visits	10.318	3.132	3.959	5.180*	-1.216	-2.404
		(9.359)	(3.505)	(3.742)	(2.832)	(2.197)	(2.704)
	Informed of complications	12.465	6.430**	10.725**	5.454*	3.394	2.046
		(8.654)	(2.972)	(4.168)	(3.113)	(2.996)	(3.603)
	Weight measurement	-2.956	-4.414	4.194	3.374*	0.241	0.692
		(7.004)	(3.012)	(2.704)	(1.981)	(1.689)	(2.086)
	Blood pressure measurement	-1.527	-5.512*	2.254	1.745	-2.168	-1.894
		(6.624)	(2.834)	(2.698)	(2.047)	(1.477)	(1.770)
	Urine sample collection	0.128	0.379	7.657**	4.369	-2.077	-4.884
		(8.645)	(3.225)	(3.796)	(2.834)	(2.502)	(3.007)
Delivery care	Institutional delivery	15.933*	7.043**	-3.661	-3.224	0.454	-0.214
		(8.166)	(3.289)	(3.096)	(2.826)	(2.138)	(2.826)
	Skilled-assisted delivery	18.138**	11.465***	-0.361	0.606	-0.106	-0.698
		(7.470)	(3.154)	(3.004)	(2.573)	(1.834)	(2.387)
Child morbidity prevalence	ARI	-7.737*	-9.677***	2.299	1.616	-0.737	-1.355
		(4.114)	(1.630)	(2.519)	(1.835)	(1.785)	(2.171)
	Fever	-8.222	-10.121***	3.107	3.297	-2.213	-3.532
		(5.991)	(2.178)	(2.893)	(2.169)	(2.107)	(2.478)
	Diarrhea	-5.054	-4.342***	0.878	-0.514	-3.221**	-4.705***
		(4.015)	(1.515)	(2.434)	(1.836)	(1.401)	(1.718)
	Obs.	1,588	958	1,531	1,088	1,422	1,026

Each *row* represents a separate regression. Clustered standard errors are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively. District-level social and economic indicators as well as regional dummies are included in all estimations.

Using the pooled sample covering the period of 1992-2014, Table 1 shows that accreditation does not have a significant positive effect on all health outcomes except the prevalence of childhood diarrhea. This unexpected finding invited us to disentangle the observed effects of the program from each time period.

For the study period 2000-2005, Table 2 provides evidence that having access to an accredited facility is associated with a higher likelihood of being informed of the side effects of contraceptives and of other methods of contraception that could be used. We find that the proportion of women with access to accredited facilities, who are informed of the side effects of contraceptives used as well as of other methods of contraception that could be used, increased significantly by 16 percentage points (ppts) and 7 ppts, respectively, compared to women with access to non-accredited facilities. These positive effects weakened during the period 2005-2008, and disappeared during the period 2008-2014.

With respect to ANC, Table 2 shows that accreditation had a limited positive effect on ANC during the period 2000-2005, specifically on being informed of signs of pregnancy complications, and no significant effect during the period 2008-2014. However, we observe multiple significant positive effects on key ANC outcomes during the period 2005-2008: the proportion of women with access to accredited facilities, who had 4+ ANC visits, increased significantly by 5 ppts compared to women with access to non-accredited facilities. We also observe significant positive effects of accreditation on being informed of signs of pregnancy complications, weight measurement, and urine sample collection during ANC visits.

In terms of delivery care, Table 2 indicates that having access to an accredited facility is associated with a higher likelihood of both institutional delivery and skilled-assisted delivery during the period 2000-2005. Institutional delivery and skilled assistance during delivery increased by more than 7 ppts and 11 ppts, respectively, among women with access to accredited facilities. Nevertheless, the estimates of both outcomes are statistically insignificant during the periods 2005-2008 and 2008-2014.

In parallel, we observe that accreditation has multiple significant positive effects on child morbidity prevalence during the period 2000-2005. Table 2 shows that accreditation reduced the prevalence of childhood ARI, childhood fever, and cjildhood diarrhea among children with access to accredited facilities by about 10 ppts, 10 ppts, and 4 ppts, respectively, compared to children with access to non-accredited facilities. We also observe a significant positive effect of accreditation on child morbidity prevalence later during the period 2008-2014, but we do not observe any significant effects on all child morbidity prevalence outcomes during the period 2005-2008.

A comparison between the early effects of accreditiaion during the study period 2000-2005 versus later during the periods 2005-2008 and 2008-2014 indicates that the positive effect of the facility accreditation program faded by time. Some of the positive effects were even reversed. One explanation could be that interventions under the HSRP have been slowing down and weakening since 2005. This trend becomes more apparent when we check the extent to which facilities comply with reform rather than the rate by which facilities join the HSRP. A plausible indicator of compliance is the accreditation score. While more facilities get accredited, accreditation scores were increasing until 2004 but started to decrease since then. Accreditation compliance also varies across governorates (Grun & Ayala, 2006). A high level of commitment to and participation in the HSRP was evident in the preparation and early implementation phases. However, successive changes in the leadership of the healthcare sector in Egypt affect the ownership of and commitment to reform efforts, undermining the reputational gains of accreditation.

Robustness tests. We test the robustness and plausibility of our results by running several alternative checks which are discussed in Appendix C. Mainly, we test the parallel-trends requirement for the acceptable application of DiD; we run placebo models; we verify the common support requirement for the feasibility of the matching; we provide tests on the quality of the matching, and do several sensitivity analyses on the matching method.

6. CONCLUSION

This paper contributes to the existing literature by investigating the effect of accreditation as a policy tool to improve quality on key patient outcomes rather than downstream process indicators. Six DHS waves are exploited to investigate the effect of quality improvement through Egypt's facility accreditation program on family planning, ANC, delivery care, and child morbidity prevalence. We use DiD fixed-effects models and also combine DiD with Kernel PSM to address the potential for endogeneity bias.

We find evidence that accreditation has multiple positive effects on family planning, ANC, delivery care, and child morbidity prevalence during the study period 2000-2005. Having access to an accredited facility is associated with a higher likelihood of being informed of the side effects of contraceptives and of other methods of contraception that could be used. Also, having access to an accredited facility is associated with a higher likelihood of both institutional delivery and skilled assistance during delivery. In parallel, accreditation is associated with lower prevalence of childhood ARI, childhood fever, and childhhod diarrhea during this period among children with access to accredited facilities.

Nevertheless, we observe that the positive effects of the facility accreditation program are not as intense during the subsequent study periods of 2005-2008 and 2008-2014, as the reputational gains of accreditation have been undermined. These results emphasize that a high level of commitment, which is a reflection of strong political will, is indispensable for the success of quality improvement interventions in low- and middle-income countries. Decentralization in no way diminishes the necessity of a high level of commitment from the central government.

The findings of this paper suggest that accreditation as a means for improving the quality of care could be associated with significant improvements in family planning, ANC, delivery care, and child morbidity prevalence in low- and middle-income countries. However, accreditation alone is not sufficient to sustain high quality of care, especially with respect to delivery care. A possible explanation is that the facility accreditation program was successful in improving the process of care provided but did not have the anticipated effect on patient outcomes. There could also be factors other than accreditation that might have affected performance differentially in accredited and non-accredited facilities. One factor is the nature and effectiveness of outreach activities carried out by facilities. Our results encourage an enquiry in this direction. Moreover, future research on this topic should broaden its scope to investigate which interventions, if combined with accreditation, could be associated with improved patient outcomes. There is evidence that improvements could be achieved, for example, through combining accreditation with properly monitored and well-designed payment or incentive schemes (Quimbo *et al.*, 2008). Nevertheless, it is important to understand the design and limitations of such an intervention, especially in resource-limited settings (Hammer & Jack, 2002).

Compliance with Ethical Standards:

Funding: The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript

Conflict of Interest: The authors declare that they have no conflict of interest.

References

Akerlof, G. (1970). The market for 'lemons:' Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3), pp.488-500.

Al Tehewy, M., Bssiouni, S., Habil, I. & El Okda, S. (2009). Evaluation of accreditation program in non-governmental organizations' health units in Egypt: Short-term outcomes. *International Journal for Quality in Health Care*, 21(3), pp.183-189.

Alkhenizan, A. & Shaw, C. (2011). Impact of accreditation on the quality of healthcare services: A systematic review of the literature. *Ann Saudi Med.*, 31(4), pp.407-416.

Arrow, K. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, 53, pp.941-969.

Brubakk, K., Vist, G.E., Bukholm, G., Barach, P. & Tjomsland, O. (2015). A systematic review of hospital accreditation: The challenges of measuring complex intervention effects. *BMC Health Services Research*, 15(1), 280.

Bukonda, N., Tavrow, P., Abdallah, H., Hoffner, K. & Tembo, J. (2002). Implementing a national hospital accreditation program: The Zambian experience. *International Journal for Quality in Health Care*, 14(Supplement 1), pp.7-16.

Cleveland, E.C., Dahn, B.T., Lincoln, T.M., Safer, M., Podesta, M. & Bradley, E. (2011). Introducing health facility accreditation in Liberia. *Global Public Health*, 6(3), pp.271-282.

Dranove, D., & Jin, G. (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature*, 48(4), pp.935-963.

El-Jardali, F., Jamal, D., Dimassi, H., Ammar, W. & Tchaghchaghian, V. (2008). The impact of hospital accreditation on quality of care: Perception of Lebanese nurses. *International Journal for Quality in Health Care*, 20(5), pp.363-371.

Flodgren, G., Gonçalves-Bradley, D.C. & Pomey, M.P. (2016). External inspection of compliance with standards for improved healthcare outcomes. *The Cochrane Database of Systematic Reviews*, 12(3).

Graham, W.J., Bell, J.S. & Bullough, C.H. (2001). Can skilled attendance at delivery reduce maternal mortality in developing countries? In: V. DeBrouwere & W. Van Lerberghe, eds. Safe motherhood strategies: A review of the evidence. *Studies in Health Services Organisation and Policy*, 17(1), pp.97-130. Antwerp: ITG Press.

Grun, R. & Ayala, J. (2006). *Impact evaluation of the Egyptian Health Sector Reform Project: Pilot phase*. Health, nutrition and population discussion paper 54512. Washington, DC: World Bank.

Hammer, J. & Jack, W. (2002). Designing incentives for rural health care providers in developing countries. *Journal of Development Economics,* 69(1), pp.297-303.

Heckman, J.J., Ichimura, H. & Todd, P. (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies*, 65(2), pp.261-294.

Hinchcliff, R., Greenfield, D., Moldovan, M., Westbrook, J.I., Pawsey, M., Mumford, V. & Braithwaite, J. (2012). Narrative synthesis of health service accreditation literature. *BMJ Quality & Safety*, 21(12), pp.979-991.

Hong, R., Mishra, V. & Fronczak, N. (2011). Impact of a quality improvement programme on family planning services in Egypt. *Eastern Mediterranean Health Journal*, 17(1), pp.4-10.

Hort, K., Djasri, H. & Utarini, A. (2013). *Regulating the quality of health care: Lessons from hospital accreditation in Australia and Indonesia*. Working paper 28. Melbourne: The Nossal Institute for Global Health.

Mason, N.M., Wineman, A., Kirimi, L. & Mather, D. (2017). The effects of Kenya's 'smarter' input subsidy programme on smallholder behaviour and incomes: Do different quasi-experimental approaches lead to the same conclusions? *Journal of Agricultural Economics,* 68(1), pp.45-69.

NCC (National Collaborating Centre for Women' s and Children' s Health), (2015). *NICE Clinical Guidelines, No. 30*. London: RCOG Press.

Oster, E. (2009). Does increased access increase equality? Gender and child health investments in India. *Journal of Development Economics,* 89(1), pp.62-76.

Pagan, A. & Ullah, A. (1999). *Nonparametric econometrics*. Cambridge: Cambridge University Press.

Pan, W. & Bai, H. (Eds.) (2015). *Propensity score analysis: Fundamentals and Developments.* New York: Guilford Press.

Quimbo, S.A., Peabody, J.W., Shimkhada, R., Woo, K. & Solon, O. (2008). Should we have confidence if a physician is accredited? A study of the relative impacts of accreditation and insurance payments on quality of care in the Philippines. *Social Science & Medicine*, 67, pp.505-510.

Rosenbaum, P.R. & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), pp.41-55.

Salmon, J.W., Heavens, J., Lombard, C. & Tavrow, P. (2003). *The impact of accreditation on the quality of hospital care: KwaZulu-Natal Province Republic of South Africa*. Bethesda: United States Agency for International Development.

Silverman, B.W. (1986). *Density estimation for statistics and data analysis*. London: Chapman & Hall.

Viswanathan, H., & Salmon, J.W. (2000). Accrediting organizations and quality improvement. *American Journal of Managed Care*, 6(10), pp.1117-1130.

Wooldridge, J.M. (2012). *Introductory econometrics: A modern approach,* 5th edition. Independence, KY: Cengage Learning.

World Health Organization (2016, 7 December). *Global health observatory (GHO) data*. Retrieved from http://www.who.int/gho/en/

Appendix A: The Facility Accreditation Program

Table A.1: Score by subarea

Quality dimension	Sub-area	Sub-area weight
Patient rights	Patient rights	2
	Dimension total	**2**
Patient care	General clinical areas	3
	Hypertension	3
	Diabetes	3
	ANC*	3
	Normal delivery, neonatal	3
	Postnatal care	3
	IMCI**	3
	Immunization	3
	Family planning	3
	Dimension total	**27**
Safety	Infection control	3
	Sterilization	3
	Employee health safety	1
	Environmental safety	2
	Dimension total	**9**
Support services	Emergency	2
	Laboratory	2
	Radiology	2
	Pharmacy	3
	Housekeeping	1
	Kitchen	1
	Laundry	1
	Dimension total	**12**
Management of information	Medical records	2
	MIS***/reporting	1
	Dimension total	**3**
Quality improvement program	Quality improvement program	2
	Dimension total	**2**
Family practice model	Prevention and screening	3
	Continuity of care	3
	Referral	3
	Dimension total	**9**
Management of the facility	Human resource development	1
	Management	1
	Budgeting	1
	Continuous education	1
	Provider satisfaction	1
	Dimension total	**5**

The scoring criteria of the accreditation standards ranges from zero to three. Scores of zero, one, two and three denote that an accreditation standard is not met, unacceptable (partially met), acceptable (partially met), and fully met, respectively. All the scores from each activity are added to get the aggregate for the accreditation standard. The average score for each standard is calculated by dividing the aggregate scores by the frequency of activities. The scores are weighed at the sub-area score level (level one) and the overall facility score level (level two) as shown in Table A.1.

*ANC: Antenatal care. **IMCI: Integrated Management of Child Illnesses. ***MIS: Management Information System.

Source: Egypt's MOH

Table A.2: Overall facility score

Quality dimension	Dimension weight	% of total score
Patient rights	1	6%
Patient care	5	29%
Safety	3	18%
Support services	2	12%
Management of information	1	6%
Quality improvement program	1	6%
Family practice model	3	18%
Management of the facility	1	6%
Total	**17**	**100%**

See notes in Table A.1.

Source: Egypt's MOH

Insert A.3: Quality dimension assessment for the different subareas

For the subarea *ANC*, the "patient care" focuses on the quality of ANC at the facility, i.e., the surveyor assesses if a comprehensive history and physical examination is performed for all patients. The general physical examination should include weight measurement, height measurement, blood pressure measurement, as well as measurement of edema of lower limbs. The surveyor also assesses if the necessary diagnostic tests (laboratory and radiology) are performed on time to determine the diagnosis. These tests include but are not limited to blood analysis, complete urine analysis, and ultrasound according to clinical guidelines. In addition, the surveyor assesses that all treatment plans are appropriate according to clinical guidelines. For example, supplementation of iron and folic acid in first trimester is checked. The surveyor also judges the number of ANC visits according to clinical guidelines and if some educational messages are discussed with patient. For example, the physician should assist pregnant women have better knowledge and understanding of their immunization status (tetanus toxoid); the importance and the number of visits prior to delivery; alarming signs such as bleeding; and, the delivery services in the facility.

The focus of "patient care" in the subarea *IMCI* is the wellbeing of children under five years of age. The surveyor assesses if a comprehensive history and physical examination is performed for all sick children according to age of child (checking for cough, diarrhea, sore throat, ear infection, and fever); if health providers explain to mothers disease classification and treatment using clear and simple language; if diagnostic tests are appropriately referred when needed; and, finally, if the facility provides appropriate prevention and treatment to all sick children according to IMCI guidelines.

The assessment of "patient care" in the sub-area *family planning* mainly focuses on the provision and quality of counseling sessions, i.e., if a comprehensive history and physical examination is performed for all new women according to guidelines; if the facility has a good IEC system such as discussing all family planning methods and the different methods, mode of action, side effects, and costs of each.

Also the *equipment* has to follow international standards in accredited facilities. Thus, if needed, accreditation is accompanied by a series of interventions so that equipment meet the expected quality standards and staff is competent in addressing family health needs. Usually, this implies upgrading, renewing, or adding modern equipment such as sterilization ovens, delivery chairs, and dentist chairs in family health units (FHU) and ensuring that there are ultrasounds and x-rays machines, and hematological and cytological labs in family health centers (FHC). To strengthen staff's competence, equipment interventions are accompanied by a comprehensive training package for facility staff. For physicians and nurses, the package focuses on family health practice. For other non-medical specialists in facilities, such as pharmacists, lab technicians and social workers, the package focuses on subject-specific training. In addition, training is a means to introduce substantial administrative changes in facilities, such as reaching out to and rostering families, and keeping medical records electronically and in family folders.

Appendix B: Descriptive Statistics

Tables B.1 highlights the difference in the district characteristics, based on which targeting took place, between accredited and non-accredited facilities. We use the two-sample t-test to check whether the means of the two groups differ significantly.

On average, districts to which accredited facilities belong have significantly higher HH overcrowding during the period 2000-2005; significantly lower illiteracy, inaccessibility to electricity, and inaccessibility to potable water, smaller family size, and bigger population size during the period 2005-2008; and significantly lower income dependency, inaccessibility to electricity, and inaccessibility to potable water, smaller family size, and lower HH overcrowding.

Thus, Table B.1 provides evidence that the actual targeting of the HSRP does not strictly follow the socio-economic vulnerability index.

Table B.1: Two-sample t-test of district characteristics of accredited and non-accredited facilities

	2000-2005			2005-2008			2008-2014		
	Non-accredited	Accredited	Difference	Non-accredited	Accredited	Difference	Non-accredited	Accredited	Difference
Illiteracy	33.093	32.069	1.023	33.300	29.912	3.387***	33.351	33.107	0.243
			(1.138)			(0.584)			(0.331)
Unemployment	9.548	9.061	0.488	9.475	9.610	-0.135	9.535	9.391	0.144
			(0.494)			(0.256)			(0.149)
Income dependency	0.591	0.596	-0.005	0.592	0.539	0.054***	0.595	0.587	0.007**
			(0.011)			(0.006)			(0.003)
Inaccessibility to electricity	1.419	0.903	0.516	1.501	0.804	0.698***	1.946	0.711	1.235***
			(0.487)			(0.261)			(0.156)
Inaccessibility to potable water	4.734	5.105	-0.372	4.900	2.450	2.451***	5.734	3.353	2.381***
			(0.863)			(0.455)			(0.268)
Family size	4.346	4.411	-0.065	4.346	4.207	0.139***	4.358	4.304	0.054***
			(0.043)			(0.022)			(0.012)
HH overcrowding	1.149	1.177	-0.028**	1.141	1.152	-0.010	1.142	1.126	0.016***
			(0.013)			(0.006)			(0.004)
Population size	31.748	29.464	2.284	31.482	33.685	-2.203**	31.521	31.176	0.345
			(1.811)			(0.931)			(0.520)

Standard errors are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Appendix C: Checks and Robustness Extensions

C.1 Parallel-trends check

The key identifying assumption of DiD is *parallel trends* in health outcomes of accredited and non-accredited health facilities in the absence of the facility accreditation program. We need to ensure that this assumption is not violated despite of two reasons: First, whether accreditation was targeted at health facilities already performing better (or worse) with respect to the health outcomes of interest. Second, the magnitude and even the sign of the DiD effect can be sensitive to the functional form if the outcomes' averages for accredited and non-accredited facilities are significantly different at the baseline. The validity of the DiD estimates depends on the treated and control units being similar at the baseline. In this section, we present a number of diagnostics we ran to assess the validity of the parallel-trends assumption.

Pre-treatment trends in health outcomes. We use information from the 1995-2000 DHS survey waves to check for parallel trends prior to the 2000-2005 period, information from the 2000-2005 DHS survey waves to check for parallel trends prior to the 2005-2008 period, and information from the 2005-2008 DHS survey waves to check for parallel trends prior to the 2008-2014 period.

Following Mason *et al.* (2017), we regress the change in health outcomes in the period 1995-2000 (i.e., pre-treatment slopes) on a dummy for if a facility is accredited in 2005 "treated" as well as facility- and district-level controls.[9] Table C.1 reports the mean changes in health outcomes between the 1995 and 2000 survey waves for facilities that are accredited versus non-accredited as of the 2005 wave, the mean changes in health outcomes between the 2000 and 2005 survey waves for facilities that are accredited versus non-accredited as of the 2008 wave, and the mean changes in our health outcomes between the

[9] In general, parallel trends is satisfied if unobserved confounding is time-invariant and additive, and becomes more plausible with pre-treatment covariates.

2005 and 2008 survey waves for facilities that are accredited versus non-accredited as of the 2014 wave.

Table C.1: Mean difference in health outcomes

	Outcome	Treated		
		1995-2000	2000-2005	2005-2008
Family planning	Knowledge of side effects	34.366	2.641	-4.965
		(20.842)	(6.995)	(5.681)
	Knowledge of contraceptives		4.468	-5.794
			(7.134)	(5.497)
ANC	4+ visits	1.367	-10.662*	0.181
		(17.757)	(6.051)	(4.735)
	Informed of complications		-10.931	1.502
			(6.906)	(5.752)
	Weight measurement		-5.453	1.541
			(5.712)	(3.367)
	Blood pressure measurement		-1.158	-2.341
			(5.447)	(3.312)
	Urine sample collection		-7.086	-0.335
			(6.666)	(4.695)
Delivery care	Institutional delivery	6.358	1.140	5.887
		(16.107)	(5.653)	(4.315)
	Skilled-assisted delivery	-4.718	0.262	8.316**
		(15.898)	(5.536)	(3.686)
Child morbidity prevalence	ARI	-13.382	1.046	1.674
		(12.820)	(4.304)	(3.017)
	Fever	-23.883	1.835	-2.552
		(16.119)	(4.942)	(3.695)
	Diarrhea	-6.025	2.826	0.416
		(9.808)	(4.217)	(3.143)

Each row represents a separate regression. The covariates are the facility characteristics, district socio-economic indicators, and regional dummies. Standard errors are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

The table indicates that the "treated" dummy is not statistically significant for all the reported health outcomes except skilled-assisted delivery in the period 2005-2008, however the estimated effect of accreditation on this particular outcome is already insignificant and negative for the period 2008-2014 (see Table 2). As the health outcomes of accredited and non-accredited facilities had moved in tandem before the facility accreditation program started, we are confident that outcomes would have continued to move in tandem in the post-intervention period. Thus, the test for pre-trends confirms that the DiD design is valid and the reported DiD estimators are unbiased.

Placebo treatment. We also run a placebo test by defining a "false" lagged accreditation intervention. If the functional form of the DiD set-up is properly specified, pre-accreditation estimations should yield null results. That is, the facility accreditation program should not have a significant effect on the health outcomes of accredited facilities before being subject to accreditation. We use the data of the study period 1995-2000 to verify the results of the period 2000-2005. For the period 1995-2000, facilities that are accredited after 2000 are defined as treated and facilities that are not accredited after 2000 are defined as control. Facilities subject to additional interventions under the HSRP are removed from the dataset. We repeat the same steps to verify the results of the periods 2005-2008 and 2008-2014.

The results of the placebo test are reported in Table C.2. The treatment estimates are not significantly different from zero for all health outcomes in 2000-2005 and for the majority of health outcomes in 2005-2008. Interestingly, several health outcomes of control facilities are significantly better than that of treated facilities for the period 2000-2005. That is, differences between accredited and non-accredited facilities reported in Table 2 only emerged after the introduction of the facility accreditation program, i.e., accreditation causes the effects observed rather than the other way around.

Placebo outcomes. Lastly, we identify some health outcomes that, theoretically, should be unaffected by the facility accreditation program, but might be indirectly. Examples of these outcomes are modern contraceptive prevalence, ANC by skilled health personnel, tetanus immunization during pregnancy, cesarean section (C-section) rates, and under-five child mortality. If the DiD design is valid, the facility accreditation program should not have any effect on the placebo health outcomes in any study period. We re-estimate the DiD model using these outcomes and report the results in Table C.3. None of the placebo outcomes are statistically significant, which supports the validity of our DiD models.

Table C.2: Difference-in-differences estimated effects of placebo accreditation

	Outcome	1995-2000	2000-2005	2005-2008
Family planning	Knowledge of side effects	0.556	5.605	3.668
		(14.402)	(4.831)	(2.660)
	Knowledge of contraceptives		9.418**	-3.791
			(4.665)	(2.662)
ANC	4+ visits	2.396	-9.143**	6.539***
		(9.444)	(4.276)	(2.330)
	Informed of complications		-1.069	0.968
			(3.965)	(2.490)
	Weight measurement		-10.219***	2.430
			(3.930)	(1.747)
	Blood pressure measurement		-3.000	2.789
			(3.906)	(1.751)
	Urine sample collection		-6.306	3.691
			(4.417)	(2.420)
Delivery care	Institutional delivery	-1.186	-2.387	0.978
		(10.150)	(4.532)	(2.462)
	Skilled-assisted delivery	-0.225	-4.243	3.418
		(10.736)	(4.307)	(2.178)
Child morbidity prevalence	ARI	-5.200	1.223	0.753
		(6.271)	(2.494)	(1.465)
	Fever	-9.331	2.886	0.662
		(7.352)	(3.083)	(1.765)
	Diarrhea	-2.337	-0.228	1.437
		(4.720)	(2.475)	(1.488)

Each row represents a separate regression. The covariates are the facility characteristics, district socio-economic indicators, and regional dummies. Standard errors are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Table C.3: Difference-in-differences estimated effects of placebo outcomes

	Outcome	2000-2005	2005-2008	2008-2014
Family planning	Modern contraceptive prevalence	5.371	-3.544	3.503**
		(5.681)	(3.055)	(1.728)
ANC	ANC by skilled health personnel	9.358	0.625	-3.817**
		(8.731)	(3.251)	(1.894)
	Tetanus immunization	-5.747	-3.615	0.063
		(7.729)	(4.247)	(2.989)
Delivery care	C-section rate	4.663	-0.017	3.366
		(6.881)	(3.739)	(2.494)
Child health	Under-five child mortality	-0.289	1.172	0.027
		(1.791)	(1.093)	(0.756)

Each row represents a separate regression. The covariates are the facility characteristics, district socio-economic indicators and regional dummies. Clustered standard errors are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

C.2 Common support check

A requirement for matching to be feasible is the common support or overlap condition. Thus, we check the overlap in the distribution of observable characteristics between treatment (accredited facilities) and control (non-accredited facilities) groups by visually inspecting the densities of propensity scores of both groups.

Figures C.1, C.2, and C3 show that there is a large common support area or a sufficient overlap in propensity scores of accredited and non-accredited facilities to produce adequate matches for all study periods. This is expected as the number of non-accredited facilities is significantly larger than that of accredited facilities. This variation also explains why the calculated propensity scores do not exceed 0.8. In principle, if there are at least as many control units as there are treated units in the data, all the treated units can be matched, but when a small caliper is used (as in this case), the matching requires that almost all the propensity scores be less than 0.5 (Pan & Bai, 2015). So, as the number of accredited versus non-accredited facilities increases in the second study period, the propensity scores increase. Plausibly, Figure C.2 shows that the control group has a higher maximum propensity score before matching, but not after matching. Figures C.2 and C.3 also provide evidence that none of the groups has a higher maximum propensity score than the other after matching.

Figure C.1: Estimated propensity scores: Kernel density estimates, 2000-2005

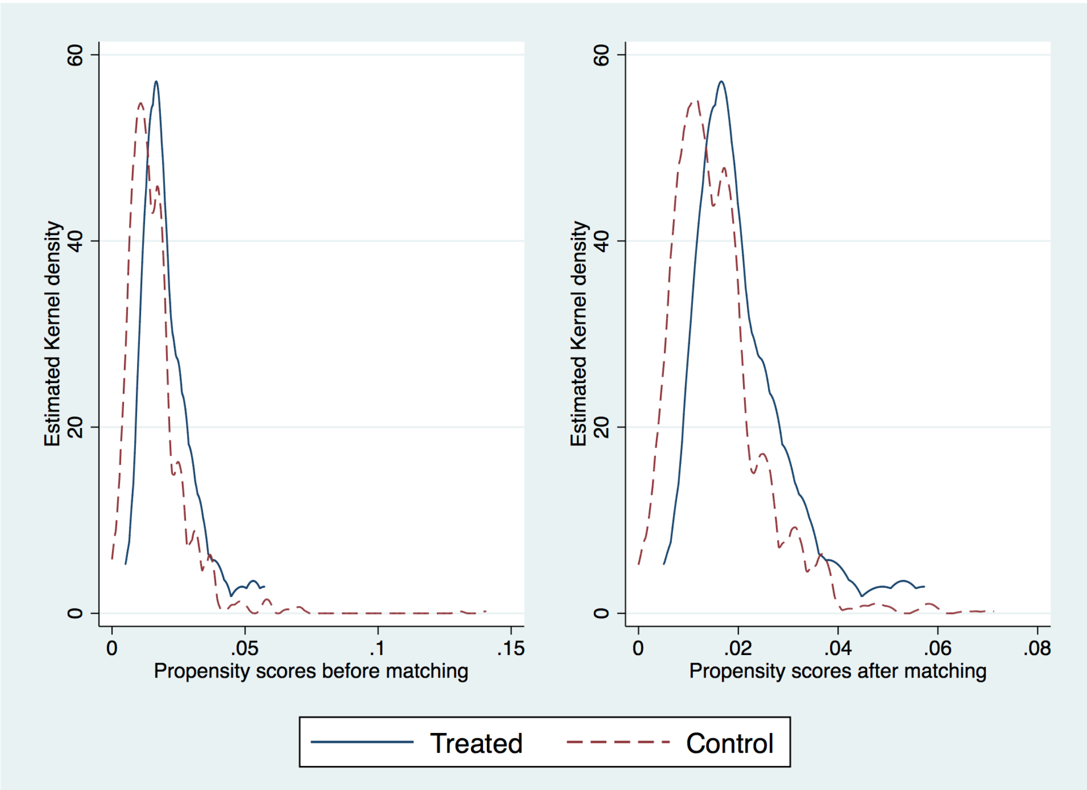


Figure C.2: Estimated propensity scores: Kernel density estimates, 2005-2008

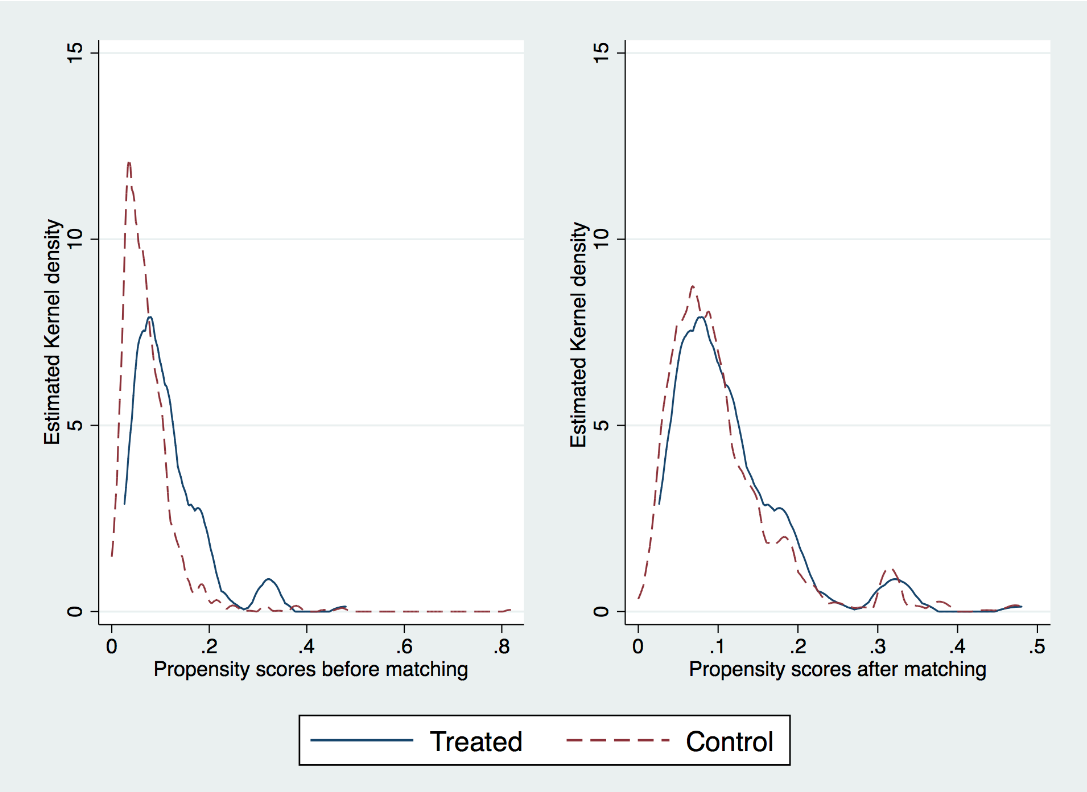
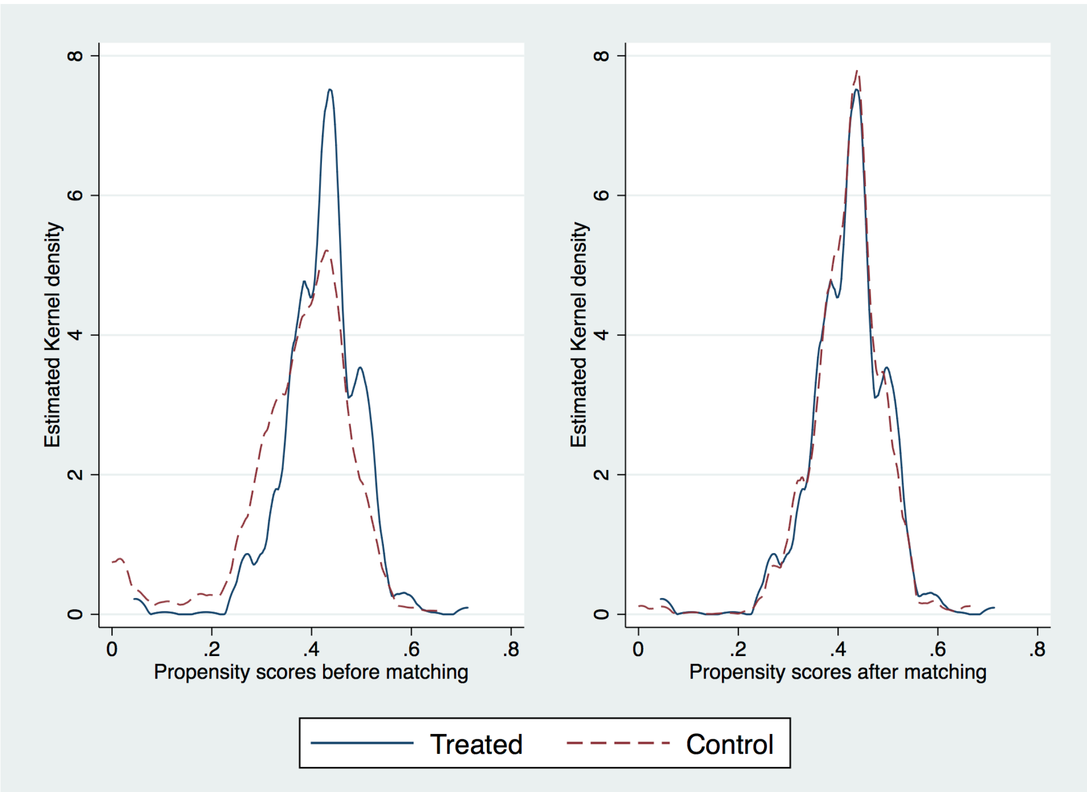


Figure C.3: Estimated propensity scores: Kernel density estimates, 2008-2014



C.3 Quality of matching

To check the extent to which observable characteristics are balanced in the matched sample, we perform the balancing t-test with the weighted covariates. Specifically, we use the balancing two-sample t-test of the difference in means of covariates across matched samples of facilities. Our covariates of interest are the ones used earlier to match treated and control health facilities. The results of the t-test are reported in Table C.4. As the table indicates, there are no systematic differences in general at the baseline in the means of observed characteristics between accredited and non-accredited facilities. That is, matching on the propensity score is successful.

C.4 Sensitivity of results

We further inspect the sensitivity of our results to the type of the Kernel function, the bandwidth of the Kernel function, and the estimation method of the propensity score. To do the Kernel matching, we must first specify the type of the Kernel function. We initially use the Epanechnikov Kernel (the default type) to obtain our main results. In Tables C.5, C.6, and C.7, we compare the main results of the estimated effects reported to the results obtained based on other types of functions, specifically Gaussian, biweight, uniform, and tricube. Overall, we find that our main estimation results are not sensitive to the type of the Kernel function.

Table C.4: Difference in mean district characteristics between accredited and non-accredited facilities

	2000-2005			2005-2008			2008-2014		
	Non-accredited	Accredited	Difference	Non-accredited	Accredited	Difference	Non-accredited	Accredited	Difference
Illiteracy	30.952	30.141	-0.811	29.142	28.819	-0.323	30.442	30.972	0.530
Unemployment	9.651	8.982	-0.669***	10.126	9.83	-0.296	10.105	9.858	-0.247
Income dependency	0.571	0.581	0.010	0.535	0.526	-0.009	0.561	0.563	0.001
Inaccessibility to electricity	1.188	0.716	-0.472**	0.663	0.645	-0.018	0.714	0.684	-0.030
Inaccessibility to potable water	3.971	3.890	-0.081	2.256	2.425	0.17	2.624	2.791	0.167
Family size	4.283	4.295	0.012	4.194	4.207	0.013	4.239	4.242	0.003
HH overcrowding	1.144	1.157	0.013*	1.153	1.154	0.001	1.131	1.135	0.003

Standard errors are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Table C.5: Sensitivity to the type of the Kernel function, 2000-2005

	Outcome	Main results	Type of function			
			Gaussian	Biweight	Uniform	Tricube
Family planning	Knowledge of side effects	15.777***	15.859***	15.721***	15.890***	15.879***
		(3.990)	(3.994)	(3.988)	(3.995)	(3.995)
	Knowledge of contraceptives	6.578*	6.636*	6.506*	6.606*	6.641*
		(3.528)	(3.534)	(3.524)	(3.537)	(3.535)
ANC	4+ visits	3.132	3.058	3.226	3.091	3.080
		(3.505)	(3.512)	(3.501)	(3.513)	(3.512)
	Informed of complications	6.430**	6.220**	6.635**	6.182**	6.235**
		(2.972)	(2.972)	(2.973)	(2.972)	(2.972)
	Weight measurement	-4.414	-4.435	-4.346	-4.376	-4.402
		(3.012)	(3.018)	(3.007)	(3.021)	(3.019)
	Blood pressure measurement	-5.512*	-5.487*	-5.492*	-5.420*	-5.462*
		(2.834)	(2.838)	(2.831)	(2.841)	(2.840)
	Urine sample collection	0.379	0.335	0.456	0.375	0.340
		(3.225)	(3.233)	(3.220)	(3.236)	(3.234)
Delivery care	Institutional delivery	7.043**	6.991**	7.086**	6.965**	6.991**
		(3.289)	(3.294)	(3.285)	(3.294)	(3.293)
	Skilled-assisted delivery	11.465***	11.502***	11.436***	11.526***	11.507***
		(3.154)	(3.159)	(3.150)	(3.160)	(3.159)
Child morbidity prevalence	ARI	-9.677***	-9.612***	-9.727***	-9.590***	-9.608***
		(1.630)	(1.634)	(1.627)	(1.635)	(1.633)
	Fever	-10.121***	-10.079***	-10.160***	-10.067***	-10.067***
		(2.178)	(2.178)	(2.178)	(2.179)	(2.179)
	Diarrhea	-4.342***	-4.232***	-4.439***	-4.176***	-4.216***
		(1.515)	(1.517)	(1.515)	(1.517)	(1.517)

Each row represents a separate regression. Standard errors are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Table C.6: Sensitivity to the type of the Kernel function, 2005-2008

	Outcome	Main results	Type of function			
			Gaussian	Biweight	Uniform	Tricube
Family planning	Knowledge of side effects	1.246	5.312	1.394	2.106	1.312
		(3.309)	(3.405)	(3.298)	(3.325)	(3.325)
	Knowledge of contraceptives	-8.643***	-5.029	-8.721***	-7.860**	-8.356***
		(3.222)	(3.325)	(3.211)	(3.234)	(3.234)
ANC	4+ visits	5.180*	5.264*	5.153*	4.806*	5.093*
		(2.832)	(2.972)	(2.811)	(2.867)	(2.856)
	Informed of complications	5.454*	6.557**	5.305*	5.889*	5.734*
		(3.113)	(3.173)	(3.103)	(3.136)	(3.128)
	Weight measurement	3.374*	4.496**	3.374*	3.613*	3.421*
		(1.981)	(2.077)	(1.971)	(2.002)	(1.993)
	Blood pressure measurement	1.745	2.721	1.731	1.869	1.766
		(2.047)	(2.145)	(2.038)	(2.071)	(2.059)
	Urine sample collection	4.369	6.270**	4.278	4.443	4.381
		(2.834)	(2.919)	(2.821)	(2.857)	(2.850)
Delivery care	Institutional delivery	-3.224	-3.663	-3.286	-3.498	-3.236
		(2.826)	(2.986)	(2.802)	(2.855)	(2.852)
	Skilled-assisted delivery	0.606	0.329	0.599	0.474	0.588
		(2.573)	(2.733)	(2.551)	(2.602)	(2.597)
Child morbidity prevalence	ARI	1.616	1.784	1.615	1.369	1.589
		(1.835)	(1.883)	(1.833)	(1.849)	(1.842)
	Fever	3.297	2.816	3.268	3.197	3.298
		(2.169)	(2.229)	(2.163)	(2.178)	(2.176)
	Diarrhea	-0.514	-0.577	-0.495	-0.449	-0.546
		(1.836)	(1.924)	(1.824)	(1.857)	(1.850)

Each row represents a separate regression. Standard errors are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Table C.7: Sensitivity to the type of the Kernel function, 2008-2014

	Outcome	Main results	Type of function			
			Gaussian	Biweight	Uniform	Tricube
Family planning	Knowledge of side effects	-3.586	-3.502	-3.654	-3.565	-3.523
		(3.465)	(3.400)	(3.461)	(3.471)	(3.468)
	Knowledge of contraceptives	0.356	-0.031	0.492	0.118	0.200
		(3.560)	(3.487)	(3.558)	(3.561)	(3.560)
ANC	4+ visits	-2.404	-2.217	-2.422	-2.400	-2.365
		(2.704)	(2.646)	(2.699)	(2.713)	(2.708)
	Informed of complications	2.046	2.375	1.866	2.185	2.259
		(3.603)	(3.528)	(3.603)	(3.604)	(3.602)
	Weight measurement	0.692	1.019	0.555	1.010	0.885
		(2.086)	(2.046)	(2.083)	(2.089)	(2.088)
	Blood pressure measurement	-1.894	-1.599	-1.959	-1.764	-1.787
		(1.770)	(1.724)	(1.768)	(1.771)	(1.771)
	Urine sample collection	-4.884	-4.667	-4.897	-4.764	-4.822
		(3.007)	(2.959)	(3.004)	(3.012)	(3.008)
Delivery care	Institutional delivery	-0.214	0.489	-0.491	0.203	0.094
		(2.826)	(2.742)	(2.826)	(2.828)	(2.824)
	Skilled-assisted delivery	-0.698	-0.218	-0.881	-0.440	-0.485
		(2.387)	(2.322)	(2.386)	(2.390)	(2.387)
Child morbidity prevalence	ARI	-1.355	-0.734	-1.447	-1.223	-1.260
		(2.171)	(2.119)	(2.170)	(2.170)	(2.171)
	Fever	-3.532	-2.851	-3.771	-3.080	-3.233
		(2.478)	(2.426)	(2.474)	(2.485)	(2.483)
	Diarrhea	-4.705***	-3.778**	-4.850***	-4.361**	-4.498***
		(1.718)	(1.661)	(1.721)	(1.710)	(1.713)

Each row represents a separate regression. Standard errors are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

To do the Kernel matching, we must also specify the bandwidth of the Kernel function. The choice of bandwidth implies a trade-off between bias and efficiency. On one hand, a small bandwidth decreases the bias of estimates as we use the most similar observations to construct the counterfactual. The characteristics of these facilities are, in general, very similar. However, a small bandwidth decreases the efficiency of estimates as we ignore a lot of information from the sample. The fact that many control facilities are not used for the estimation implies an increase in the imprecision of estimates caused by a higher variance. On the other hand, a large bandwidth increases both the bias and efficiency of estimates. The bandwidth choice is, therefore, a compromise between a small variance and an unbiased estimate of the true density function. This choice is more important in practice than the choice of the type of the Kernel function (e.g., Silverman, 1986; Pagan & Ullah, 1999).

The default bandwidth of the Kernel function initially used to obtain our main results is 0.06. Alternative bandwidths are tried (bandwidths = 0.05 and 0.1). Table C.8 shows our main results of the estimated effects using different bandwidths. We find that our main results are not sensitive in general to the bandwidth parameter.

The estimation of propensity scores depends on a parametric specification (commonly logit or probit), which affects the quality of matching and, consequently, the results. As for the benchmark we use a probit model, we test the results when we use a logit model instead and then re-run the PSM DiD models. The results of this exercise are reported in Table C.9. We find that the estimates for both methods of estimation match for most outcomes.

The previous robustness checks rule out an existing trend that could challenge the PSM DiD identifying assumptions. Our robustness checks also provide evidence that the main estimation results reported in section 5 are not sensitive in general to alternative types of the Kernel function, bandwidths of the Kernel function, and estimation methods of the propensity score.

Table C.8: Sensitivity to the bandwidth of the Kernel function

	Outcome	2000-2005			2005-2008			2008-2014		
		Main results	Bandwidth 0.05	Bandwidth 0.1	Main results	Bandwidth 0.05	Bandwidth 0.1	Main results	Bandwidth 0.05	Bandwidth 0.1
Family planning	Knowledge of side effects	15.777***	15.777***	15.869***	1.246	1.868	3.766	-3.586	-3.683	-3.655
		(3.990)	(3.989)	(3.995)	(3.309)	(3.285)	(3.380)	(3.465)	(3.457)	(3.404)
	Knowledge of contraceptives	6.578*	6.393*	6.640*	-8.643***	-8.546***	-6.504**	0.356	0.689	-0.074
		(3.528)	(3.524)	(3.535)	(3.222)	(3.197)	(3.291)	(3.560)	(3.555)	(3.491)
ANC	4+ visits	3.132	3.375	3.051	5.180*	4.985*	4.902*	-2.404	-2.425	-2.349
		(3.505)	(3.500)	(3.512)	(2.832)	(2.789)	(2.927)	(2.704)	(2.694)	(2.654)
	Informed of complications	6.430**	6.872**	6.203**	5.454*	5.142*	5.977*	2.046	1.612	2.164
		(2.972)	(2.974)	(2.972)	(3.113)	(3.094)	(3.157)	(3.603)	(3.604)	(3.533)
	Weight measurement	-4.414	-4.089	-4.431	3.374*	3.381*	4.128**	0.692	0.464	0.847
		(3.012)	(3.008)	(3.019)	(1.981)	(1.960)	(2.047)	(2.086)	(2.080)	(2.049)
	Blood pressure measurement	-5.512*	-5.321*	-5.481*	1.745	1.712	2.288	-1.894	-1.998	-1.731
		(2.834)	(2.835)	(2.839)	(2.047)	(2.029)	(2.114)	(1.770)	(1.764)	(1.730)
	Urine sample collection	0.379	0.587	0.331	4.369	4.007	5.394*	-4.884	-4.834	-4.812
		(3.225)	(3.222)	(3.234)	(2.834)	(2.806)	(2.899)	(3.007)	(3.001)	(2.958)
Delivery care	Institutional delivery	7.043**	7.079**	6.984**	-3.224	-3.451	-3.666	-0.214	-0.819	0.417
		(3.289)	(3.282)	(3.294)	(2.826)	(2.772)	(2.933)	(2.826)	(2.825)	(2.755)
	Skilled-assisted delivery	11.465***	11.421***	11.504***	0.606	0.535	0.291	-0.698	-1.110	-0.293
		(3.154)	(3.149)	(3.160)	(2.573)	(2.525)	(2.682)	(2.387)	(2.384)	(2.331)
Child morbidity prevalence	ARI	-9.677***	-9.699***	-9.603***	1.616	1.620	1.442	-1.355	-1.588	-0.955
		(1.630)	(1.624)	(1.634)	(1.835)	(1.835)	(1.864)	(2.171)	(2.170)	(2.122)
	Fever	-10.121***	-10.120***	-10.072***	3.297	3.199	2.883	-3.532	-4.062	-2.984
		(2.178)	(2.178)	(2.178)	(2.169)	(2.157)	(2.209)	(2.478)	(2.470)	(2.430)
	Diarrhea	-4.342***	-4.449***	-4.219***	-0.514	-0.502	-0.523	-4.705***	-5.020***	-3.974**
		(1.515)	(1.515)	(1.517)	(1.836)	(1.812)	(1.895)	(1.718)	(1.722)	(1.667)

Each row represents a separate regression. Standard errors are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Table C.9: Sensitivity to the estimation method of the propensity score

	Outcome	2000-2005		2005-2008		2008-2014	
		Probit	Logit	Probit	Logit	Probit	Logit
Family planning	Knowledge of side effects	15.777***	15.758***	1.246	2.180	-3.586	-3.540
		(3.990)	(3.994)	(3.309)	(3.310)	(3.465)	(3.462)
	Knowledge of contraceptives	6.578*	6.513*	-8.643***	-7.883**	0.356	0.466
		(3.528)	(3.531)	(3.222)	(3.226)	(3.560)	(3.558)
ANC	4+ visits	3.132	3.128	5.180*	4.912*	-2.404	-2.451
		(3.505)	(3.506)	(2.832)	(2.836)	(2.704)	(2.701)
	Informed of complications	6.430**	6.520**	5.454*	5.467*	2.046	1.899
		(2.972)	(2.973)	(3.113)	(3.108)	(3.603)	(3.600)
	Weight measurement	-4.414	-4.297	3.374*	3.388*	0.692	0.638
		(3.012)	(3.014)	(1.981)	(1.987)	(2.086)	(2.085)
	Blood pressure measurement	-5.512*	-5.435*	1.745	1.832	-1.894	-1.902
		(2.834)	(2.836)	(2.047)	(2.052)	(1.770)	(1.771)
	Urine sample collection	0.379	0.394	4.369	4.401	-4.884	-4.882
		(3.225)	(3.227)	(2.834)	(2.836)	(3.007)	(3.005)
Delivery care	Institutional delivery	7.043**	6.948**	-3.224	-3.558	-0.214	-0.268
		(3.289)	(3.288)	(2.826)	(2.829)	(2.826)	(2.827)
	Skilled-assisted delivery	11.465***	11.395***	0.606	0.457	-0.698	-0.741
		(3.154)	(3.154)	(2.573)	(2.584)	(2.387)	(2.388)
Child morbidity prevalence	ARI	-9.677***	-9.631***	1.616	1.351	-1.355	-1.479
		(1.630)	(1.629)	(1.835)	(1.832)	(2.171)	(2.168)
	Fever	-10.121***	-10.015***	3.297	3.055	-3.532	-3.674
		(2.178)	(2.179)	(2.169)	(2.168)	(2.478)	(2.475)
	Diarrhea	-4.342***	-4.283***	-0.514	-0.474	-4.705***	-4.777***
		(1.515)	(1.516)	(1.836)	(1.840)	(1.718)	(1.716)

Each row represents a separate regression. Standard errors are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.