When a Doctor Falls from the Sky:
The Impact of Easing Physician Supply Constraints on Mortality

Edward N. Okeke* Isa S. Abubakar†

Abstract

This paper studies the effect of easing physician supply constraints in a large developing country. In this unique policy experiment, conducted in coordination with the Nigerian government, physicians were randomly assigned to primary health service areas. The physicians were posted to government health centers serving these areas. Prior to the arrival of the physicians, health care in the facility was provided entirely by mid-level health care providers. To separate skill from volume effects, another group of service areas was provided with an additional mid-level health worker. We find that the arrival of the physicians led to a significant reduction in mortality for newborns. Our results imply a 3-4 percentage-point reduction in mortality among infants whose care was provided, at least in part, by a physician. Using the estimated value of lifetime earnings, we calculate that the physician program generated nearly $7 million dollars in value (or about $1.7 million dollars in net present terms). Comparing this to the cost of the intervention, we estimate that each $1 spent on the program returned nearly $8 in benefits.

*Department of Economics, Sociology and Statistics, RAND, 1200 South Hayes, Arlington, VA 22202. Email: eokeke@rand.org.

†Department of Community Medicine, Bayero University Kano and Aminu Kano Teaching Hospital.

This project would not have been possible without the help of many people: We thank the Director-General and State Directors of the NYSC program without whose help this project could never have been carried out. We are grateful to the Commissioners of Health of Akwa Ibom, Bauchi, Gombe, Jigawa, and Kano states whose support was vital. We thank staff of the Primary Health Care Development Agency and State Ministries of Health for their help in implementing the project. We thank Nnenna Ihebuzor, Ibrahim Yisa, Christie Akwaowo, and Usman Bashir whose support was invaluable at various stages. We thank Obinna Onwujekwe and the staff of the Health Policy Research Group, University of Nigeria for providing research and administrative support. We are grateful for the dedicated work of Bamidele Aderibigbe, Saidu Abubakar, Sadia Aliyu, Sadiya Awala, Yakubu Suleiman, and Edidiong Umoh who oversaw the field work. We thank Susan Lovejoy, Onyinye Stephen-Gow, and Laura Pavlock-Albright for project management support, and Juliana Chen, Crystal Huang, Stephen Okpalaononuju, Adeyemi Okunogbe, and Victor Olajide for excellent research assistance. We thank Peter Glick and A.V. Chari, who were involved during the early stages of this project, for many helpful comments and discussions. Finally, we are grateful to all the study participants, who gave generously of their time. This study was funded by the Eunice Kennedy Shriver National Institute for Child Health and Human Development through Grant No: R01HD083444. The findings, interpretations and conclusions expressed in this article are those of the authors. The funder played no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

1 Introduction

Ensuring equitable access to health care is a fundamental policy objective (Healthy People 2020).[1] However in many countries skilled medical professionals are in short supply, making it difficult to achieve health policy goals (World Health Organization, 2014, 2016). Estimates of the supply deficit for physicians range from 0.5 million to nearly 3 million (Liu et al., 2016; World Health Organization, 2014). Physician labor supply constraints are particularly acute in developing countries which account for an increasingly large share of the global disease burden.[2] This is not just a poor country issue (see European Commission, 2012; Ishikawa et al., 2013; Australia, 2012).[3] That this is a first-order policy issue can hardly be in doubt (Grobler et al., 2015; World Health Organization, 2016).[4] In poor countries, the scarcity of physician labor is believed to be a key binding constraint on improving health outcomes (Scheffler et al., 2009; Liu et al., 2016).[5] However, at least one recent paper has raised questions about whether expanding access to physicians improves health (Carrillo and Feres, 2019). Other studies have raised similar questions (Laurant et al., 2018).

This paper provides new evidence on the effect of easing physician supply constraints in a large developing country. We present results from a first of its kind policy experiment that took place in Nigeria. In this experiment, conducted in coordination with the government, certain communities were randomly selected to receive a physician. This physician was posted to the public health center serving the community, where they worked for approximately one year. Prior to the arrival of the physician, health care in the facility was provided entirely by mid-level health care workers.[6] By posting a physician to the health center, we effectively displaced care that would otherwise have been provided by a mid-level health provider, allowing us to study the effect. To separate skill effects (the effect of adding a physician) from volume effects (the effect of adding another health care worker), another group of communities was randomly selected to contemporaneously receive an additional mid-level health provider (similar to those in the health facility). A third group of communities, that received no additional health care workers – or other inputs – served as a control group. To examine the impact of adding a physician on outcomes, we recruited more than 10,000 pregnant women residing in these communities, tracking their outcomes through delivery, and the outcomes of their children into early infancy.[7]

[1]Much of the attention usually focuses on the demand side of access (Finkelstein et al., 2012; Kolstad and Kowalski, 2012; Miller and Wherry, 2017; Sommers et al., 2012; Shigeoka, 2014; Goodman-Bacon, 2018). However, the supply side of access, while it does not get as much attention, is at least as important, as the recent COVID-19 pandemic has demonstrated.

[2]Nearly all maternal and child deaths, for example, occur in poor countries (Bhutta et al., 2014).

[3]A quarter of the US population resides in a federally designated primary care health professional shortage area (Health Resources and Services Administration, 2020). Scholars have warned about a looming shortage of primary care physicians (Petterson et al., 2012; Cooper et al., 2002; Assocation of American Medical Colleges (AAMC), 2020). A spate of recent media articles have highlighted this issue, raising alarm about the likely effects of diminished access on health outcomes. See for example: https://n.pr/2FFBy2h, https://wapo.st/2Td897E, https://lat.ms/3hPLcQx, and https://wapo.st/36FqoGI.

[4]Programs such as the Conrad-30 Visa Waiver program for foreign-trained physicians and the National Health Services Loan Repayment Program, in the US, are examples of public policies designed to address physician supply. Similar programs exist in other countries (Fontes et al., 2018).

[5]These constraints are likely to bind even tighter when there are large-scale external shocks such as a disease epidemic.

[6]This includes nurses, midwives, community health workers, and other non-physician health care providers.

[7]We focused on pregnant women and young children because this was a population with a high probability of health care utilization during the intervention period.

We document a number of key findings: first, we find that (slightly) loosening physician supply constraints has large and significant impacts on health outcomes. In intent-to-treat models, the probability that a child born in our sample died in early infancy decreased by more than 20 percent. On average, we estimate reductions of about 6-8 infant deaths per 1,000 live births. This is not a trivial effect: it would eliminate the newborn mortality gap between rural and urban areas in Nigeria, and would close about half of the gap between the poorest households (bottom 20%) and the richest (the top 20%) (National Population Commission and ICF International, 2014). These effects are even more remarkable considering that the program supplied just one additional physician to each treated community for less than a year. We show that the results are robust to a range of checks.

To explain these effects, we present compelling evidence that the introduction of physicians led to a major upgrade in quality. Physicians demonstrated higher levels of clinical ability at baseline, performing better on multiple assessments (this result is to be expected). The skill distributions are quite distinct: a 20th percentile physician performs about the same as an 80th percentile mid-level provider. Interestingly, we find that (length of) practice experience by mid-level health providers—capturing the effect of on-the-job learning—does not impact the size of the skill gap: experienced mid-level providers in our sample performed similar to early stage mid-level providers. Drawing on rich survey and observational data, we show that the skill gap translates into observable (and measurable) differences in quality. This result is based on observations of thousands of provider-patient interactions, follow-up surveys of women, and data collected from clinic managers who were asked to carry out confidential evaluations of the new health providers.

We then move beyond intent-to-treat analysis to estimate local average treatment effects. The question we ask is: if the same patient were to be treated by a physician vs. a mid-level health provider, how would their outcomes differ? The randomized assignment of physicians gives us a way to gain traction on this question. However, it does not take us all the way, because even though the provider is randomized, patients are not. To get around this, we exploit another randomized intervention that was nested within the supply experiment. About half of the women in the sample were randomly selected to be offered a conditional cash incentive—a cash payment conditioned on attending prenatal visits, receiving facility care at childbirth and postnatal care. By leveraging this external shock to demand, combined with the randomized assignment of physicians, we are able to provide a robust answer to this question. We find large returns to physician human capital in this context: early newborn mortality decreases by an additional 3-4 percentage points when health care is provided, at least in part, by a physician (instead of a mid-level provider). We find that the incremental health gains of seeing a physician are almost as large as the main effect of formal care (relative to informal or no care) when it is provided by mid-level providers.

The results in this paper speak to an ongoing debate on the substitutability of physicians and non-physicians in health care production (Carrillo and Feres, 2019; Kleiner et al., 2016; Traczynski and Udalova, 2018; Markowitz et al., 2017).[8] Our results suggest that this depends crucially on who and what physicians

[8]The stakes are high: $564 billion was spent paying for physician services in the US in 2018, or approximately 16% of total spending (Rama, 2020). If non-physicians deliver similar outcomes as physicians less expensively, as some have argued (see for example Perloff et al., 2016), this has the potential to reduce health care spending.

are substituting for. In this context we find that non-physicians are very poor substitutes for physicians.[9] These findings are likely to generalize to other low and middle-income country settings. These results have implications for policies that propose to address the scarcity of physician labor in poor countries by shifting more tasks to mid-level health care workers (McPake and Mensah, 2008; The Lancet, 2018; World Health Organization, 2014). This paper illumines the trade-offs. Having lower-level health workers performing tasks that they may not be adequately trained to perform has implications for patient welfare.[10]

In the concluding part of the paper, we present a cost-benefit calculation. We estimate that between 89 and 208 lives were saved by the physician program. Using the estimated value of lifetime earnings, we calculate that the physician program generated nearly $7 million dollars in value (or about $1.7 million dollars in net present terms). Comparing this to the cost of the intervention, we estimate that each $1 spent on the program returned nearly $8 in benefits. Inflating costs in line with a scaled-up version of the program, we estimate that physicians would return about $3 in benefits for every dollar spent. Physicians are expensive to produce (Mills et al., 2011; Scheffler et al., 2009), but our findings demonstrate a large potential return on investment. Over a period of a year, we estimate that each physician would conservatively return about $35,000 in net present value. Multiplied over a full career, which might last 30-40 years, it is clear that a physician would return, many times over, what it would cost to produce them.[11]

The constraint on physician supply often does not lie with the demand for medical education, but with training capacity.[12] However, unlike in rich countries such as the US where physicians must complete several more years of training after graduating from medical school (the medical residency) before they can be licensed, and where the number of residency slots creates a bottleneck (Assocation of American Medical Colleges (AAMC), 2019; Goodman and Robertson, 2013), physicians in many developing countries can practice straight out of medical school (usually after completing a mandatory year of supervised practice), meaning that it is comparatively simpler to expand supply. One additional point: history has taught us that expanding physician supply is necessary but not sufficient for improving access in underserved communities. Combining this with deliberate distributive policies will also be important.[13]

The results in this paper make novel contributions to our understanding of the returns to human capital in health care delivery (Bartel et al., 2014; Doyle et al., 2010). Health care accounts for a significant portion of GDP in many countries, and this share is increasing over time (Kotlikoff and Hagist, 2005; Papanicolas et al., 2018). In the US, medical spending as a share of GDP increased from 13.3% in 1996 to 18% in 2016

[9]The study includes various types of non-physicians, so these are not a narrow set of conclusions.

[10]This feeds into some of the debates about occupational licensing in health care and its merits (Kleiner, 2016; Anderson et al., 2020). The argument for licensing regulation is that it helps to protect consumers. This paper shows there is some merit to this idea. Our results imply that non-physician health providers deliver worse care that translates to higher rates of avoidable deaths.

[11]Obviously, as supply expands, marginal costs will likely increase (though equilibrium wages may also fall) and marginal benefits will decrease, but the available evidence suggests that we are long way away from the point of diminishing returns.

[12]According to the Joint Admissions and Matriculation Board (JAMB) which handles admissions to Nigerian universities and colleges, about 1.65 million students applied for admission in 2018 (in 2020, applications exceeded 1.9 million). More than 400,000 applied to a program in the Faculty of Medicine, Pharmacy and Health Sciences. Only 29,715 were admitted in 2018 (National Bureau of Statistics, 2018). Nigeria has 37 fully accredited medical schools and 5 with partial accreditation, with spaces for about 4,000 students (Medical and Dental Council of Nigeria, 2019). By comparison, the US graduated about 26,000 students from medical schools in 2018 (Kaiser Family Foundation, 2019). The size of the college age population (18-24 years) in both countries is not that different: 31 million in the US (de Brey et al., 2018) compared to about 28 million in Nigeria based on estimates.

[13]Filling vacancies in underserved areas using physicians signed to temporary renewable contracts is one model. See Bärnighausen and Bloom (2009) for a review of other strategies.

(Dieleman et al., 2020). Health care production requires a complex mix of inputs, of which labor is a key component, and an understanding of the relationship between human capital and productivity in health care has enormous implications for how health care is organized and delivered (Baicker and Chandra, 2018; Chandra and Skinner, 2012). We also make a contribution to the literature on the returns to health care in the formal sector in developing countries (Adhvaryu and Nyshadham, 2015; Godlonton and Okeke, 2016). We show that the returns to the use of formal health care intuitively depends on the quality of that care (relative to the outside option).

The rest of the paper is structured as follows: in Section 2 we describe the experiment and provide relevant institutional details; in Section 3 we describe the data and our sample; in Section 4 we describe the analysis and present our findings; concluding remarks are in Section 5.

2 Institutional Details

2.1 Policy context

With an estimated population of more than 200 million, Nigeria is the 7th largest country in the world. Gross National Income per capita is about $2,000 (World Bank, 2020). Nigeria scores poorly on most welfare indices: more women die in Nigeria each year during pregnancy and childbirth than in any other single country in the world (World Health Organization, 2019b); about 10% of all newborn deaths globally occur in Nigeria (Lawn et al., 2014); the under-five child mortality rate is 132 per 1,000 live births, worse than that of much poorer countries (National Population Commission and ICF International, 2019), and the average life expectancy is 54 years (compared to 61 years on average for sub-Saharan Africa). There are marked rural-urban disparities in health outcomes—the infant mortality rate in urban areas is 65 per 1,000 compared to 88 per 1,000 in rural areas—as well as geographic disparities, with states in the north-east and north-west exhibiting generally worse health indicators than states in the south-east and south-west (National Population Commission and ICF International, 2019). These poor outcomes reflect, in part, disparities in access to health care resources.

To properly understand the context it is important to have an idea of how the Nigerian health care system is organized. Nigeria operates a tiered health care system with primary health centers forming the base of the pyramid. Primary health centers provide a set of services defined by national guidelines that include general outpatient care, maternal and newborn care, nutrition, control of communicable diseases, non-communicable disease prevention, and health education (National Primary Health Care Development Agency, 2014). Many primary health clinics also provide inpatient care. About 80% of Nigeria's approximately 30,000 primary health care facilities are in the public sector. Responsibility for primary health care in Nigeria is shared between the National Primary Health Care Development Agency, which sets guidelines and policy, and local governments which manage primary health care centers. Secondary hospitals serve as referral centers for patients needing more advanced medical care (though patients can also seek care there directly). University Teaching and Specialist hospitals occupy the apex of the health care pyramid. Most medical care in Nigeria is paid for out-of-pocket, though some types of medical care, such as maternal and child health care, are subsidized in the public sector (Okonofua et al., 2011).

Available data suggests that Nigeria has a physician per capita ratio of about 0.4 per 1,000 people (World Bank Data Bank, 2020).[14] This falls short of the minimum number recommended by the World Health Organization: 1 per 1,000 people (World Health Organization, 2014). For comparison, the United States and United Kingdom have 3 and 2.8 physicians per 1,000 people respectively (Young et al., 2019; Moberly, 2017). As is the case in many other countries, physicians tend to be concentrated in urban centers. They are usually found in secondary and tertiary health facilities, and many primary health centers are staffed by mid-level health providers. In a preliminary survey that we conducted, 3 in 4 primary health centers reported that they had never had a physician on staff. These issues are recognized by the Nigerian government, which has tried to address them in various ways (see for example Okeke et al., 2016).

2.2 Participating sites

180 primary health service areas or HSAs (hereafter clusters) were selected to take part in the policy experiment. A HSA consists of communities served by a common government primary health center. Service area boundaries are administratively drawn. Each HSA in our sample covers about 7,000 people so that, collectively, the 180 HSAs that participated in the experiment cover nearly 1.3 million individuals. The government health center in the HSA is the main source of care for community residents; 4 out of 5 facility births in our sample, for example, are in this health center. To provide a mental picture of what they look like: they are small to medium-sized facilities, ranging in size from 1 to 56 beds, with an average of 14 beds. They have an average of five health care providers on staff, ranging from nurses to community health workers. About 4 in 5 provide round-the-clock, 24-hour, health care services.[15]

The participating HSAs were selected from five states representing three of Nigeria's six geopolitical regions. Two states were selected from the north-west region (Kano and Jigawa), two from the north-east (Gombe and Bauchi), and one from the south-south region (Akwa Ibom). A map is provided in Figure A.1. As noted before, the north-west and north-east have some of the worst outcomes in Nigeria. While official statistics are hard to come by, many states in these regions have a shortage of physicians. Often the only physicians are located in the general hospital in the nearest large town. The fifth state in the south was included with an eye towards external validity. The south is richer, better educated, and has comparatively more health resources per capita. We chose a state from the south-south region because this is the worst-performing of the three southern regions. The specific states in each region were chosen in consultation with our local partners. Feasibility of implementation was a key consideration.

36 HSA clusters were chosen from each state; 180 in total. The clusters were chosen with the help of government health officials in each state. The clusters were drawn from underserved areas of the state—which is true of most parts of the state outside of the state capital—and the health centers serving the cluster had to offer pregnancy and delivery services. The clusters were broadly distributed across the state both for representativeness and to minimize crossover between clusters. Health officials began with a comprehensive list of government primary health centers located in underserved local government areas

[14] Data from the Medical and Dental Council of Nigeria suggests that the actual number of practicing physicians is closer to 0.2.

[15] The remainder also provide service after hours because a health worker lives on the premises, or lives nearby and can be summoned when needed.

and narrowed this down based on the considerations above. Across all five states, about 80% of local government areas (or LGAs) are included.[16] 50% of LGAs have one cluster, 31% have two clusters, 15% have three clusters, and 5% have four clusters. While not strictly a random sample, the clusters should be considered broadly representative of each state. The list of 180 clusters was finalized in the fall of 2016. To avoid confusion, moving forward whenever we use the term 'site' or 'cluster', we are referring to the primary health service area.

2.3 Experimental Design

The 180 clusters were randomly assigned to one of three arms: added physician, added mid-level health provider, and control. There are 60 clusters in each arm. See Figure 1a. Within each state, sites were first grouped into blocks (or strata), with randomization carried out within each block. Large LGAs with up to 3 sites were treated as a single block, adjoining LGAs with less than three sites were combined into blocks by grouping adjacent areas. We chose this strategy primarily because administrative responsibility for primary health centers is at the local government level, making it administratively easier to manage implementation. Additionally, since communities in the same (or adjacent) local government area share common characteristics this helps to increase precision. In total we created 43 blocks. Blocks range in size from three to nine clusters. In the analysis that follows, we include block fixed effects in our regression models. Randomization was carried out on a computer.

A secondary experimental intervention was nested within the trial (see Figure 1b). 50% of census areas in each HSA were randomly assigned to a conditional cash incentive intervention in which households with a pregnant woman were offered a cash payment of N5,000—approximately $14 at the prevailing exchange rate—to be made after the birth of the child if the pregnant woman attended at least three prenatal visits, gave birth in a health care facility, and attended a postnatal visit. All three conditions had to be met; there were no partial or pro-rated payments. $14 is equivalent to about 30% of monthly household food expenditures (Nigerian National Bureau of Statistics, 2016). In the remaining 50% of census areas, households were not offered the incentive, but participating women received small gifts worth about $0.43 at endline to thank them for participating. The results of the cash incentive intervention are described in Okeke and Abubakar (2020). In our analysis of the effects of the health provider intervention, we include a dummy indicating whether the woman was offered a conditional incentive. We also test whether there were interactions between both interventions (there were not). Later on, we will also leverage the external shock to demand provided by the conditional incentives.

The sample of participating pregnant women and their infants born during the intervention period constitute the primary source of data for this study (enrollment is described in Section 3.1.1). We chose to focus on pregnant women and young children for two reasons: first, because this was a population that had a high probability of needing/using medical care during the intervention period and, second, because child health is sensitive to the quality of health care, particularly care provided early in life (Lawn et al., 2014; Sankar et al., 2016).

[16] An LGA is similar to a US county.

Ethical approval: Ethical approval for the study was given by RAND's Human Subjects Protection Committee and by the Ethics Committee of Aminu Kano Teaching Hospital, Nigeria. We also sought and received approvals from all of the participating state governments. In the next section, we provide additional details about the implementation of the supply intervention.

2.4 Project implementation

Clusters randomly assigned to the physician and mid-level health provider arms each received one additional health care provider—a physician and mid-level health provider respectively—that was posted to the primary health center serving the cluster. Clusters in the control arm received no additional health care providers. No other inputs were provided. Throughout the rest of the paper, we will refer to clusters randomly assigned a physician, a mid-level provider (MLP), or nothing as physician clusters, MLP clusters, and control clusters respectively.

2.4.1 Physicians

The physicians posted to physician clusters came from a government program known as the National Youth Service Corps (NYSC). The NYSC is a one-year community service program that all graduates of Nigerian colleges and universities, with only a few exceptions, are required to participate in.[17] Even those who attend college or university in another country have to take part if they wish to eventually work in Nigeria.[18] Physicians take part after they have completed one year of post-graduate medical training (other professions usually take part immediately after graduating). Program participants receive what are known as 'call-up letters' at the beginning of their community service year. These letters indicate the states where they have been posted to carry out their community service and when they are to report to state orientation camps.[19] The postings are done centrally by the NYSC head office.

The service year officially starts with a three-week long orientation camp in the state of assignment, after which participants receive their specific work assignments within the state. A similar length of time is reserved at the end of the program for exit formalities, so even though the program nominally lasts for one year, only about ten of these months are actually spent working.[20] Postings to work sites in the state are usually done by the State NYSC office. Physicians are typically posted to government health facilities, though in large cities some might also be posted to the private sector (depending on availability). Program participants receive a modest monthly stipend from the Federal government—at the time of the study about $60 per month—but employers can also pay additional stipends. Physicians, specifically, typically receive additional allowances from their employers, but even with these allowances, the program remains a cheap

[17] The only exceptions are individuals older than 30 at graduation or those who have previously served in the Nigerian Military or Police Force. More details about the program can be found at `https://www.nysc.gov.ng`.

[18] A certificate of completion (or an exemption certificate) is usually requested by employers. This has led to a few high-profile scandals. See for example: `http://bitly.ws/8W2Z`

[19] A goal of the NYSC program is to foster national unity and a feature is to post graduates to states other than their states of origin or residence

[20] One day each week is also reserved for community service.

source of physician labor.[21] As a result, there is intense lobbying for physicians.[22] Program participants generally have no input in the process and have to report wherever they are posted. They can request to be redeployed, and these requests are handled on a case-by-case basis and may or may not be granted. The program provides a very rare instance in which physicians do not get to choose where they locate and work.

With the help of NYSC officials, physicians were posted to primary health centers in physician clusters. Logistics were arranged with the Primary Health Care Development Agency. In addition to their statutory allowance, physicians posted to a participating site received an additional monthly allowance of $143 (in one state it was bumped up to $200). The physicians were deployed between February and September 2017 (see Figure 2). Physicians were successfully posted to 58 (of the 60) clusters. In one cluster, the physician was only present for one month before redeploying so that, effectively, 57 sites received a physician. Of these, in 4 sites the assigned physicians requested redeployment and were replaced by the program.[23] The primary analysis is handled on an intent-to-treat (ITT) basis, so treatment is defined based on assignment, not whether a physician was successfully posted to the cluster. 83% of physicians were male, their average age was 28.4 years, and they had 2.1 years of clinical experience on average.

2.4.2 Mid-level providers

The posting of mid-level health care providers was handled by state governments. Nearly all were new hires; a few were redeployed from another health facility outside of the sample. The mid-level providers are of two types: community health officers (CHOs) and community health extension workers (CHEWs). These are the modal type of health care provider in primary health centers (see Table 1). They are licensed primary health care professionals whose training, licensure and practice is regulated by the Community Health Practitioners' Registration Board of Nigeria (Ordinioha and Onyenaporo, 2010). Junior CHEWs receive two years of training and are awarded a Certificate in Community Health on completion. After a few years of work experience, they can go through a three-year training program to become CHEWs.[24] On completion, they are awarded a Diploma in Community Health. CHEWs can also go on to become Community Health Officers. The CHO training program takes two years and successful trainees are awarded a Higher Diploma in Community Health.

Mid-level health providers were successfully posted to all 60 sites. 19% of the posted workers were CHOs, 76% were CHEWs, and 5% were Junior CHEWs. 63% were female, their average age was 28.7 years, and they had 2.4 years of clinical experience on average. Similar to the physicians, their postings lasted about a year. They were paid salaries ranging from $114 to $143 monthly depending on the prevailing wage in the state. We note that short-term postings for health care providers in this settings are not uncommon: another skilled worker program, the Nigerian Midwives Service Scheme, signed midwives to one-year

[21] An analogy is to American professional sports where rookie wages are fixed below market rates.

[22] The allocation process is opaque and it is probably safe to assume that workplace assignments are not random.

[23] In one case they cited personal reasons, and in the other three they redeployed for health-related reasons.

[24] They receive training in anatomy and physiology, medical sociology, psychology, pharmacology, microbiology, reproductive health, child health, nutrition, and environmental health, in addition to clinical training and supervised clinical experience. A copy of the curriculum can be found here: `https://bit.ly/33IRFYg`.

contracts (Okeke et al., 2017). The resumption dates for mid-level providers in each block was aligned with that of the physicians. As Figure 2 shows, this was largely successful.

3 Data

3.1 Sample and data sources

The analysis draws on multiple datasets. The primary source of data on individual outcomes is an in-home survey administered to women enrolled in the study (more on this below). We combine this with administrative data on health provider start and end dates, surveys of primary health center managers and health providers, and data from unannounced visits to health centers (also referred to as audit visits). Each data source will be described in more detail below.

3.1.1 Women's survey

At baseline, we enrolled approximately 60 pregnant women in each cluster into the study. Sampling was carried out by randomly drawing census areas from the cluster—the sampling frame was provided by the National Population Commission—and then enrolling all eligible pregnant women into the sample. All eligible pregnant women in a census area, who gave consent, were enrolled (only seven women, in total, declined to participate). Women in their first or second trimester of pregnancy were eligible to be enrolled. We excluded late trimester women to maximize the length of exposure. To identify pregnant women in each area, our research assistants carried out a census, going house-to-house. Pregnancy was based on self-reporting.[25] These enrollment visits took place between March and August 2017.[26] We wanted to enroll women around the time when the providers were expected to start in order to maximize pregnancy exposure. If we recruited too early, then too many women might deliver before the provider arrived. Conversely, if we recruited too late, then women might be delivering after the provider had completed their posting. Figure A.2 shows how enrollment overlapped with the arrival of the new provider posted to the cluster.

On enrollment, each woman completed a baseline intake survey that collected demographic information about her and her household. It also collected detailed information about the woman's birth history. These surveys were administered by trained research assistants using computer tablets. Women received another home visit approximately three months after giving birth. At this follow-up visit, they completed a follow-up survey. In cases where the mother was deceased, another adult household member—typically the surviving spouse—was interviewed. The survey collected information about health care utilization,

[25] We considered pregnancy tests but ultimately decided against it because of costs and the ethical implications of asking women to undergo pregnancy tests at our behest. This means that women who were not aware of being pregnant at the time are, by definition, not included in the sample. It also introduces the possibility of misreporting, but there is no reason to expect misreporting to be different by treatment arm. This was more of a concern for the conditional incentive intervention: women who were not pregnant might claim to be pregnant in order to be enrolled (with the hope of getting pregnant later). Okeke and Abubakar (2020), under the assumption that misreporting, if present, would be more likely among early-stage pregnancies, finds little evidence of misreporting. A similar proportion of women reported an early-stage pregnancy in incentive and non-incentive areas.

[26] The incentive program was announced during the same visits.

including whether the woman received care from a physician during the pregnancy or delivery. The outcome of the pregnancy and the survival status of the child was also recorded. If a child had died, the date of death and age at death was recorded. Weight and length of surviving children was measured by the research assistant. On completing the follow-up survey, a woman's participation in the study ended. Endline visits took place between September 2017 and August 2018.

Health cards: Women, at registration for prenatal care, normally receive maternal health cards where visits are recorded. Women retain these cards, and bring them along whenever they visit the health facility. Each woman enrolled in the study was provided with a card. The name and designation of the provider seen during each visit is recorded on the card. During the follow-up visit, we asked to see these cards and recorded whether the woman saw a physician during the pregnancy or delivery (physicians in Nigeria always use their designation (Dr.) so this was easy enough to identify). The main limitation with these cards is that they are not available for all women (they are available about 65% of the time). The probability of having a card at endline is similar in each arm: 65% of women in physician clusters had a card at endline, compared to 67% and 64% in MLP and control clusters, respectively.

Anecdotally, women who received care outside of the health center reported that providers were less likely to complete the cards.[27] To the extent that some of these women who received care in another health facility received care from a physician, the cards will underestimate the actual prevalence of physician-provided care. However, only 8% of women in the sample who used formal health care, did so in a government or private hospital, where a physician might be available. As long as this proportion is similar across experimental arms, which it is, then we will be undercounting in the same way across groups and our estimates of the between-group differences in means will be correct. A related issue is that women sometimes forgot to take their cards with them to the facility, or the provider omitted to record the utilization. This can be seen from Figure A.3: 1 in 10 women who gave birth in the health center did not have this recorded on the card.[28] This could also lead to an undercount. However, as Figure A.3 clearly shows, this kind of under-reporting is also not differential between arms, and so does not pose a threat to internal validity. In the analysis that follows, we will use the card as our primary source of data on the prevalence of physician-provided care, but we will also present results using women's self-reports from the follow-up survey.

3.1.2 Health centers

There are 180 primary health centers in the sample (one per cluster). We visited each primary health center twice: once shortly after the new health care provider was expected to have arrived, and again shortly before the provider's tenure ended. During these visits we surveyed the health center manager or 'in-charge' (senior health care providers who have the added responsibility of managing the operation of the health center) to collect data about the health facility. We also surveyed one health care provider in

[27] This is borne out by the data: of the 140 women who gave birth in a government hospital only 25 had this recorded on the card.

[28] In general, for women for whom we have a card, there is 89% agreement between what is reported in the survey and what is recorded on the card.

each health center.[29] Additionally, in health centers that received a new health care provider, we also surveyed that provider. The provider surveys collected demographic information, clinical qualifications and experience. It also included modules designed to assess their level of clinical ability, using a combination of multiple-choice questions, case studies, and patient vignettes. Providers were assessed in three areas: basic medical knowledge (this was tested using ten multiple choice questions); management of obstetric and newborn emergencies (this was tested using two case studies of postpartum hemorrhage and a non-breathing newborn), and management of outpatient primary care conditions (this was tested using two patient vignettes: tuberculosis and pediatric malaria).[30] These modules were administered by medically trained personnel on the research team.

We also observed each of these providers as they provided care to patients. The medically trained observer sat in a corner of the consultation room where they could observe the interaction but did not otherwise interfere with the consultation. For each interaction, the observer noted if any physical examinations were carried out, and whether a diagnosis was provided. They also noted the quality of communication, e.g., whether the health care provider explained the diagnosis and treatment in common language to the patient. For patients presenting with common symptoms such as fever, they recorded whether the provider adhered to recommended protocols. We have some information about the treatment, e.g., whether the patient was prescribed antibiotics or an injection (intramuscular or intravenous). We also have a small amount of information about each patient, collected while they were sitting in the waiting area. We know their age and sex, mode of transportation to the clinic, illness severity (rated on a scale from 0-10), and health status (poor, fair, good, very good or excellent).

3.1.3 Audits

During the intervention period, we made several unannounced visits to the health centers. During each visit, we recorded whether the health center was open, and whether the posted health care provider was present in the health center. These data give us an objective measure of the availability of the new health care providers. Health centers received an average of just over three audit visits over the intervention period.

3.2 Attrition

Attrition is not a major concern in this study. None of the 180 health centers dropped out; dropout was also negligible in the women's sample. We enrolled 10,852 pregnant women at baseline, 10,699 (98.6%) were successfully re-contacted at endline. Of these, 113 refused consent for the endline (41 in the control and physician arms, and 31 in the MLP arm). The overall attrition rate was 2.45%: 2.9% in control clusters, 2% in MLP clusters, and 2.4% in physician clusters (p-value from joint test = 0.39). Table A.1 examines

[29] We usually tried to interview the provider that saw the most cases.

[30] In the case studies, we presented a brief introduction about the case and asked a series of questions, for example, about what they would do given a set of specified clinical findings. The correct responses/actions were pre-specified and visible to the interviewer who recorded which, if any, were mentioned. The patient vignettes were similar, but longer and more detailed, requiring the subject to systematically proceed through history taking questions and physical examinations, to diagnosis and treatment. We scored their responses afterwards.

determinants of attrition. We find that women without formal schooling, women with more prior births, and women who were not offered the conditional incentive, were more likely to have dropped out. We can reject the null that attriters had similar characteristics to non-attriters. In Table A.2 we compare the baseline characteristics of women who dropped out in each arm to test whether there was a differential pattern of attrition by experimental arm. We find no evidence of this, and we fail to reject the null. The p-value from an omnibus test is 0.99.

3.3 Main outcome

The primary outcome is early newborn mortality. Nearly half of all deaths among children under five years occur in the first month of life; 75% of these deaths happen within the first week, making this a critically important time in the life of a child (Liu et al., 2015). Mortality soon after birth is known to largely be a function of events surrounding childbirth, and is often used as a marker of the quality of intrapartum care (Gabrysch et al., 2019; United Nations Children's Fund, 2019; World Health Organization, 2019a). The main causes of newborn infant deaths: complications of prematurity, intrapartum complications, and infections are all largely preventable (Lawn et al., 2014; Liu et al., 2015). We define two main outcomes: early newborn mortality (death of a liveborn infant between Day 0—the day of birth—and Day 6), and very early newborn mortality (death of a liveborn infant within the first 24 hours—deaths between Day 0 and Day 1). We will refer to these throughout the paper as 7-day and 1-day mortality, respectively. We follow the standard medical definition in defining these as deaths per 1,000 live births, but we will also present unconditional results where we define the outcome as newborn deaths per 1,000 women. The results are not sensitive to definition.

3.4 Sample definitions

For clarity, there are 10,586 women for whom we have endline data. 9,410 of these women carried their pregnancy to term and gave birth, 1,176 did not (including 19 women who died before giving birth). Intermediate outcomes, such as whether a woman received care from a physician during pregnancy or childbirth, will be analyzed at the level of the individual woman. Health outcomes such as mortality will primarily be analyzed at the level of the individual child. There were 9,126 live births in our sample (395 infants were stillborn). The primary mortality outcomes are defined as deaths per 1,000 live births (so N = 9,126).

We will also examine secondary mortality outcomes that use different cuts of the data, depending on their definition. For example, perinatal mortality, which is defined as a stillbirth or early newborn death, includes all births—live births and stillbirths (so N = 9,126 + 395 = 9,521). In utero mortality, which is defined as a miscarriage or stillbirth, includes all births + pregnancy losses (so N = 9,521 + 1,176 = 10,697). Similarly, overall child survival, which is defined as the probability that a child that was in utero at enrollment was alive at endline, also includes all births + pregnancy losses (so N = 9,521 + 1,176 = 10,697).

4 Analysis and Results

4.1 Summary Statistics and Balance Tests

Table 1 provides summary statistics. Columns 2-4 report baseline variable means by treatment arm. The variables are grouped into three panels: Panel A reports means of health service delivery variables derived from the health center survey, Panel B reports means of women's characteristics derived from the baseline survey, and Panel C reports means of fixed child characteristics, such as the sex of the infant, derived from the endline survey.

The experimental framework relies on the random assignment of study participants. To evaluate the validity of this assumption, we compare variable means across arms. We regress the variable shown in each row on the treatment indicators and strata dummies (to account for the blocked design). Standard errors are adjusted for clustering at the HSA level—reflecting both clustering in the sampling and clustering in assignment (Abadie et al., 2017). We report p-values from tests of equality between all pairwise combinations—MLP vs. control, physician vs. control, and physician vs. MLP—in Columns 4-7. In Column 8, we report p-values from a test of joint equality across all three arms for each variable. At the bottom of each panel we report p-values from an omnibus test for all the variables in the panel. The null is that none of the variables are predictive of treatment assignment, i.e., they are all jointly equal to zero.

One can see that randomization was successful and characteristics are balanced. The number of statistically significant differences are about what one would expect due to chance. This is confirmed by the joint tests. We fail to reject the null that the variables in Panel A (p = 0.19), Panel B (p = 0.18), or Panel C (p = 0.88) do not predict treatment assignment.

4.2 Did the intervention increase provider supply?

The intervention, if successful, should have had two effects: (1) increasing the number of health care providers in MLP and physician clusters, and (2) increasing the number of physicians in physician clusters. It is important to show this, because it helps to confirm that the health care providers actually arrived at their assigned health centers and, importantly, continued to be there throughout their posting. We also want to confirm that there was no offsetting withdrawal of existing providers. First, we want to confirm that there was a net increase in the number of health care workers in the primary health centers serving MLP and physician clusters. We examine this visually in Figure 3.

We graph the average number of health care providers on the staff register in the health center prior to the arrival of the new health care worker or T0, at the start of the intervention or T1, and just before the end of the intervention or T2. The intervention period officially begins with the arrival of the new health care worker and ends with their departure. As noted before, we visited health centers just after the health care provider's scheduled arrival date (T1), and again just before their expected departure (T2). The number of workers prior to the arrival of the newly deployed provider is the count of health care workers on the staff register, excluding the new provider. Given random assignment, we expect the average number of health care providers in each arm, prior to the arrival of the newly assigned worker, to be the same. With the arrival of the new health care provider, the number of health care providers in the treatment arms should increase by approximately one and, assuming no differential changes, this net increase should continue to be present towards the end of the intervention. As Figure 3 shows, this is almost exactly the case. There's a small net increase, at T2, in the number of health workers in control clusters, but we also see a similar increase in physician clusters.

Next, in Figure 4 we examine whether there was a net increase in the number of physicians in health centers in physician clusters. We examine the probability that there was a physician on the staff register at the same time points; T0, T1, and T2. We compare means across the three arms. We can see that there were almost no physicians working in the health centers prior to the arrival of the new health care workers: only two health centers—one in a control cluster and another in a physician cluster—had a physician on staff at baseline. We see the expected sharp divergence at T1, in physician clusters only, that persists until T2. The probability is less than one because not all health centers that were supposed to get a physician actually got one.

In Figure 5 we use the audit data to show that the assigned health care providers were present throughout their tenure, not only at the beginning and at the end. We estimate the probability that the posted provider was present in the health center during each surprise visit over the duration of their posting (Month 0 is the start month of the posting and Month 10 is the end month). We present smoothed local polynomial regression lines and 95% confidence intervals. The lines trend downwards over time indicating that the assigned providers were more likely to be present towards the beginning of their tenure than towards the end, but, overall, they show that the new health care providers were largely present. On average, the new physicians were less likely to be in the health center than the new mid-level providers. One potential explanation is that physicians were more likely to live further away from the health center: 88% of new mid-level providers lived either in the same community as the health center or a neighboring community, compared to 70% of physicians.

Based on the graphical evidence, we conclude that the intervention had the anticipated effects on the supply of health care providers in the cluster. Having shown these visually, we confirm these results using regressions. We estimate linear regressions of each outcome on the treatment assignment indicators. We estimate the following model:

$$Y_{kst} = \alpha + \beta_{1p} \sum_p Provider_{ks} + \beta_2 T_t + \beta_{3p} \sum_p Provider_{ks} * T_t + \phi_s + \epsilon_{kst} \tag{1}$$

Y denotes the outcome of interest in the health center serving cluster k in strata s at time t. t indicates whether the observation is from the start of the intervention period (T_1) or from the end (T_2). The unit of observation is the health center (N = 180). $\sum_p Provider$ are the treatment assignment indicators that denote whether the health center was in a control cluster, an MLP cluster, or a physician cluster. ϕ_s are strata fixed effects. Standard errors are adjusted for clustering at the health center level.

The results are in Table 2. Columns 1-2 present the results for the number of health care providers in the health center, with and without strata dummies, and Columns 3-4 present the results for the probability that a physician was available in the health center, also with and without strata dummies. The results confirm what the figures have shown. Columns 1 and 2 indicate that in MLP and physician clusters at

T1—just after the provider should have arrived—there was approximately one more health care provider on staff at the health center (0.92 and 0.79 more health providers respectively). We cannot reject the null that the coefficients are equal (p-values are reported at the bottom of the table). By T2, there was a net increase of 0.4 health care workers in control clusters. However, we cannot reject the null of a similar increase in physician clusters (p = 0.78). In MLP clusters the p-value is significant at the 10% level (p = 0.07). Columns 3 and 4 show that the probability that a physician was available in the health center at T1 was 94% in physician clusters. This probability did not change between T1 and T2 and there was also no change in MLP and control clusters. Next, we examine the effects on physician-provided medical care.

4.3 Intent-to-treat effect on physician-provided medical care

While we have shown that the physicians arrived, and were available, in the health centers serving the project clusters—a necessary first-step—an as-yet unanswered question is how this affected care provision. In particular, how did it affect the probability that a patient's medical care was provided by a physician? There are several reasons why this is important to examine. First, as Figure 5 has shown, physicians were not always present in the health centers. Second, even when physicians were on duty, it does not automatically mean that they were working hard (Das and Hammer, 2014). One can imagine physicians being present but only attending to a few patients before going home. Third, even if physicians were on duty in the health center and working hard, this does not mean that they would see all the patients coming in for care. Cases would naturally be shared among all health care providers on duty (health centers had, on average, five other health care providers). It therefore makes it an open question how the physician supply intervention translated into physician-provided medical care.

Again, we begin with a graphical treatment. We focus on two outcomes: (i) the probability that a woman saw a physician at least once during prenatal care, and (ii) the probability that the birth of the child was attended by a physician. We examine this using, both, information from women's health cards and women's self-reports from the endline survey. The advantage of the survey data is that we have it for all women. The limitation is that reporting may be subject to some degree of measurement error. One can imagine that a woman might not always know whether she was being seen by a physician or a mid-level provider. Type II errors—women reporting that they saw a mid-level provider when, in fact, they saw a physician—are likely to be more common given that physicians are relatively scarce in these communities (and women know this). This suggests that physician-provided care in physician clusters will likely be under-reported in the survey. This would bias treatment effect downwards. As such, the survey estimates provide a lower bound.

One way to assess the signal-to-noise ratio in the survey data is to examine women's responses when we know, with certainty, whether there was a physician present in the facility where care was provided. As a starting point we examine the responses of women who gave birth in the health centers. For women giving birth in health centers in control and MLP clusters, where we know that there were no physicians available, less than 1% of women—0.4% and 0.7% respectively—reported a physician present at the birth. In contrast, in health centers in physician clusters, where we know that there was a physician present, nearly 10% of women reported that a physician attended the birth. As noted before, this is likely to be an underestimate. Illustrating this, when asked at endline whether there was a physician available in the health center, only 48% of women in physician clusters replied “Yes”. In other words, over half of women in physician clusters did not know that a physician was available there.[31] According to the health card data, about 15% of births were attended to by a physician suggesting that the underreporting bias may be about 5 percentage points.[32]

In Figure 6, we graph the mean probability that: (i) a woman saw a physician at least once during prenatal care (Figure 6a), and (ii) the delivery was attended by a physician (Figure 6b). The results are broken out by treatment arm. The figures on the left use information from health cards, while the figures on the right use information reported in the endline survey. As we expected, given the likelihood of underreporting, the estimated treatment effects using data from the survey are smaller but, regardless of the source of information, the conclusions are the same. The data show that the physician treatment significantly increased the probability that a woman’s care was provided by a physician during her pregnancy and delivery. To nail down these results, we turn to regressions. The estimation model is a linear model with the following specification:

$$Y_{iks} = \alpha + \beta_{1p} \sum_{p} Provider_{ks} + \gamma X_{iks} T_t + \epsilon_{iks} \tag{2}$$

The models are estimated on an intent-to-treat basis. Y denotes the outcome of interest for woman i in cluster k in strata s. X is a vector of control variables that includes the woman’s age, ethnicity, religion, literacy level (based on whether she could read a simple sentence shown to her in English), a measure of pregnancy risk (based on a prior history of a stillbirth), a dummy indicating whether she was offered a conditional cash incentive, decision-making authority in the household (a dummy denoting whether the spouse is the sole decision-maker), a household asset index (the number of assets owned by the household), and strata dummies. The standard errors are adjusted for clustering at the HSA level.

The results are in Table 3. The results in Table 3a are based on the health cards; in Table 3b we use women’s self-reports from the endline survey. We report both unadjusted and covariate-adjusted results. The control group means are reported at the bottom of the table. Our preferred estimates in Panel A indicate that women in physician clusters were about 21 percentage points more likely to have received care from a physician during pregnancy, and about 7 percentage points more likely to have received care from a physician during childbirth. The estimated treatment effects based on women’s self-reports are smaller: about 7 percentage points for physician care during pregnancy and 3 percentage points for physician care during delivery. For reasons that we have discussed earlier, the survey-based estimates are likely to be biased downwards. Accounting for underreporting closes the gap between the two estimates. However, all of the results are highly statistically significant. Relative to means in the control arm, they imply increases of between 150 and 170 percent. In both samples, adjusting for covariates makes no difference to

[31] When we restrict to women who used the health center for prenatal or delivery care, who should have better information, this increases to 62%. However, this means that 1 in 3 women in physician clusters *who used the health center* did not know that there was a physician working there.

[32] As another validity check, we look at differences in the probability that a physician was reported present by delivery location. We expect probabilities to be much higher in private health facilities and in government hospitals. This is exactly what the data show (see Figure A.4).

the estimates.

In thinking about the size of this effect, it is important to realize that about 53% of women in the sample gave birth at home, and 25% of women did not use any prenatal care. As a proportion of women who actually used health care services, the effect is larger. Additionally, since this is an intent-to-treat estimate, it includes all physician-assigned clusters independent of whether they actually got a physician. Back-of-the-envelope calculations show that these estimates have face validity. According to our data, physicians worked about 30 hours a week. Assuming that births are smoothly distributed throughout time, this would suggest that about 18% of births in the health center would have taken place when the physician was present, i.e., 30 hours/(7 days $\times$ 24 hours). This is very close to the estimated effect based on the health card data—about 15% of births were attended by a physician in health centers where there was, in fact, a physician. Having shown that the physician treatment significantly increased the probability that care was received from a physician, we examine what effect this had on health outcomes.

4.4 Intent-to-treat effect on infant mortality

Our primary outcomes are 1-day and 7 day infant mortality. We begin with some descriptive analysis. In Figure 7 we graph means and 95% confidence intervals for each mortality outcome by treatment arm. The estimates are from a regression of each outcome on the treatment assignment dummies and randomization strata. These graphs provide the first indication of a decrease in mortality in physician clusters. In Figure 8 we unpack this some more by looking at how mortality trends evolved over time in each arm. We plot average 1 and 7-day mortality rates by quarter of birth. We aggregate by quarter for stability of the estimates (mortality is a relatively rare outcome). We can clearly see a divergence in mortality rates over time: it decreases in physician clusters while increasing slightly in the other two arms. Descriptively at least the data provide evidence that in clusters randomly assigned a physician, newborn infant mortality rates fell. To put this result on a firmer footing, we again turn to regressions. The primary value of regression estimation is that it allows us to include controls that soak up additional variation and increase statistical precision.

The estimation model is the same as in Equation (2) except that i now denotes the individual child. The results are in Table 4. In Table 4a the outcome is 1-day mortality, and in Table 4b it is 7-day mortality. The results in the first column in each table are adjusted only for stratification, the second column flexibly controls for seasonal trends in mortality (we allow trends to vary by state), and the third column includes additional individual-level controls. In addition to controlling for the mother's characteristics, we also control for birth characteristics that are known to be correlated with mortality. We include indicators for first births, male infants, and a multiple delivery (Vogel et al., 2013; Kupka et al., 2009; Astolfi and Zonta, 1999). In the first three columns in each table, the denominator is live births. In Columns 4 and 6, the denominator is all women, and the interpretation of the coefficients is as deaths per 1,000 women. The coefficients are scaled so that they can be directly interpreted as deaths per 1,000. The standard errors are adjusted for clustering.

The regression results confirm the descriptive results. They indicate that 1-day and 7-day mortality decreased in physician clusters by between 6-8 deaths per thousand live births. The unconditional results are similar. In general the results look quite similar across specifications. Relative to means in the control group, this represents a 28 and 21 percent decrease, respectively, in 1-day and 7-day mortality. To put this mortality reduction into context, it is equivalent to the total decrease in newborn mortality in Nigeria over the three decades between 1990 and 2018 (National Population Commission and ICF International, 2019). We are not able to reject the null of a similar decrease in 7-day mortality in physician and MLP clusters, but for 1-day mortality, in some of the specifications, we can reject this at the 10% level.

4.4.1 Treatment dosage/Duration of exposure

Women, and by extension the children in utero, were exposed to the treatment for different lengths of time. Some women, for example, gave birth in the same month that the physician arrived, others gave birth several months after. This matters because the later the birth relative to when the physician arrived, the longer the period of potential exposure to the treatment. This is because women had more time within which to attend prenatal visits and, mechanically, the odds of seeing the physician increases with the number of visits. More prenatal visits is also associated with a greater likelihood of a health facility delivery (see Figure A.5), and thus of having a physician present at the birth. Also, given that health workers were being posted to new communities, which is particularly important for the physicians nearly all of whom were relocating from out-of-state, one can imagine that there would be an initial adjustment period during which new health care providers might have seen fewer patients than they would once they had settled in. Both of these imply that women who gave birth further into the tenure of the physician—women with longer duration of exposure—would have been more likely to see the physician. Or to put it in terms of dosage, they received a greater dose of the treatment.

Figure 9 shows this. We examine how the probability of physician-provided medical care varies by exposure duration, calculated as the number of months between when the provider arrived and when the pregnancy ended. By our definition, a woman who gave birth a month after the provider arrived was exposed for 1 month. Women who gave birth before the arrival of the new health care provider (or in the same month) are coded as having zero months of exposure. The maximum possible length of exposure is 10 months. The median is four months.[33] In Figure 9 we plot smoothed local polynomial regression lines and 95% confidence intervals. We can see that length of exposure matters, and suggests that we should disaggregate health effects by length of exposure.

We take a first pass at this by graphing 1-day and 7-day mortality rates by the number of months of exposure in Figure 10. We see from the figure that mortality rates in physician and MLP clusters are pretty similar for children born to mothers who were exposed for a short length of time to the intervention. However, with longer exposure, one can see that mortality rates start to diverge. To firm up this result, we again turn to regression estimation. To give us more power, we divide the sample into a 'low dose' sample (less than the median length of exposure) and a 'high dose' sample (greater than the median). We then re-estimate the mortality results for each of these samples (this corresponds to a fully interacted model). Results are in Table 5. Table 5a shows the effect on 1-day mortality, and Table 5b shows the effect on 7-day

[33] The full distribution is shown in Figure A.6. The peak at zero in the distribution for physician clusters is because, in clusters that were assigned a physician but did not receive one, all women are coded as having zero months of exposure.

mortality. We present the same specifications shown in Table 4. The comparison is between physician and MLP clusters.

The results indicate that the overall decrease in newborn infant mortality is driven by women that were exposed for longer to the physician treatment (i.e., they received a greater dose of the treatment). The estimated decrease in 1-day mortality in the high dosage sample in physician clusters is 9-12 deaths per thousand, while for 7-day mortality, it is 10-13 deaths per thousand. To put these results into perspective, if newborn mortality rates in Nigeria as a whole were to reduce by a similar margin, this would move Nigeria up 31 places on the global newborn mortality rankings, from 188 (out of 194 countries) to 157 (World Bank Data Bank, 2020). For reference, Afghanistan and Somalia, both war-ravaged countries, were ranked 190 and 191 respectively. The US, with a newborn mortality rate of 3.5 per 1000 live births was ranked 46. We find no measurable effects for women who received a low dose of the treatment. In all cases, the point estimates are small and statistically insignificant.[34]

4.5 Effect on secondary health outcomes

In Table 6 we examine a range of other mortality outcomes. We examine the effects on the probability of an in utero child death (a miscarriage or stillbirth), a perinatal death (a stillbirth or an early newborn death), a newborn death (death of a liveborn infant between day 0 and day 28), and the probability of overall child survival (the probability that a child who was in utero at enrollment was alive at the endline). We disaggregate the results by treatment dosage. We find no effects on the probability of an in utero child death, but we find significant reductions in perinatal deaths (a reduction of 20 deaths per thousand), newborn deaths (a reduction of 16 per thousand), and an increase in the probability of overall child survival (increased by 3 percentage points), for women that received a higher dose of the treatment.

Better medical care might also improve child health on other dimensions. Quality prenatal care has, for example, been associated with increased birthweight (Evans and Lien, 2005; Gonzalez and Kumar, 2018; Gajate-Garrido, 2013). We do not have birthweight data for all children in the sample because majority of births took place outside of formal health facilities. As a result, only 23% of the children in our sample were weighed at birth. However, there is no difference across experimental arms in the likelihood of being weighed at birth: 22% of children in the control arm were weighed at birth, compared to 22.4% in the mid-level provider arm, and 23.7% in the physician arm (p = 0.45). For the subset of children for whom we have birthweight data, we can examine whether children born in physician clusters weighed more at birth. We can also test for differences in the incidence of low birthweight infants (infants weighing less

[34]There is an additional check we can run to assess the veracity of these results; also in the spirit of a dose-response test. Intuitively, the places where physician-provided care increased the most, should also be the places where we find the largest mortality effects. The states are a natural level of aggregation given that the sample was stratified by state. We create an indicator for whether a woman in the sample received care from a physician during her pregnancy or delivery, and disaggregate the results by state. The results are presented in Figure A.7. We plot the treatment effect estimates from state-level regressions along with their 95% confidence intervals. Panels A and B show results using the health card and survey data respectively. Both show the same results. Two states, Gombe and Jigawa, stand out in terms of the magnitude of the first stage effect. In Table A.3, we examine the mortality effect in these two states compared to the other three states that saw smaller increases. Consistent with the results in Figure A.7, we find large and significant intention-to-treat effects on both 1-day and 7-day mortality in Gombe and Jigawa states (a reduction of 9-10 deaths per thousand live births) and a smaller effect, on average, in the other three states (a reduction of 5-6 deaths per thousand live births). The latter does not reach statistical significance.

than 2.5 kilograms at birth). Low birth weight is believed to contribute, directly or indirectly, to as much as 60% to 80% of all neonatal deaths (Lawn et al., 2005, 2014; Watkins et al., 2016). The results are reported in Table A.4. The coefficients on birthweight are positive but do not reach statistical significance. We find suggestive evidence of a decrease in the incidence of low birth weight in physician clusters, but we are hesitant about drawing any firm conclusions given that we only have data for a quarter of the sample.

We also examine effects on child weight and length at endline. During the in-home follow-up survey, which was conducted, on average, three months after birth, research assistants measured weight and recumbent length for all surviving children. We have data for 99% of surviving infants in our sample. In our models we take the natural log of these variables to account for skewness in their distribution. Overall, we find no measurable effects on either weight or length. These results are reported in the Appendix. See Table A.5.

4.6 Sub-group analysis

In this section we examine whether the treatment had different effects for various sub-groups. We caution that the study was not powered for subgroup analysis and, as such, we generally cannot reject the null that the coefficients in sub-groups are the same. The results of the sub-group analysis are reported in Table 7. We focus on the primary mortality outcomes. In Columns 1-2, we examine whether there were differential ITT effects for boys vs. girls. Male infants have a well-documented mortality disadvantage (Naeye et al., 1971). In our data, for example, 7-day mortality for male infants was 1.3 percentage points higher than for female infants. A reasonable question to ask is whether boys might reap larger benefits from access to physician-provided care. Indeed, our data shows that the treatment effects were larger for males than for females. We find similar results for both 1-day and 7-day mortality.

In Columns 3-4, we examine whether there were differential effects by maternal literacy. As noted in the data section, at baseline we showed women a simple sentence in English and asked them to read it out loud. We disaggregate the effects for illiterate women—those who could not read at all—vs. women who could read the sentence in full or in part. Literacy is obviously strongly correlated with the quantity of schooling: 90% of women who could not read reported no formal schooling (Islamic schooling for purposes of this analysis is considered informal schooling) compared to 1% of women who could read the whole sentence; but it also captures the quality of the education (61% and 15%, respectively, of women with primary and secondary education could not read the sentence). Better educated mothers might be more likely to benefit from access to a physician. They might, for example, exhibit more agency and insist on seeing the physician when experiencing problems; they may also be better able to comprehend and follow instructions and advice. The results lend some support to this hypothesis. More literate women experienced larger decreases in early newborn mortality. This is quite noticeable for 7-day mortality where we see that mortality decreased by about 20 deaths per thousand among infants born to somewhat literate mothers.

In Columns 5-6, we examine whether there were differential effects by maternal pregnancy risk. The risk characteristics that we considered, all of which are known to be associated with higher risk of an adverse birth outcome, were: first-time births, mother's age (mothers less than 18 or older than 35 years are considered to be higher risk), history of vaginal bleeding during the pregnancy, and a history of a prior miscarriage or stillbirth. We compared mothers with none of these risk factors to women with at least one of the risk factors. We caution that this is, at best, a crude measure of risk. It is not the case that women without any of these risk factors have zero risk, but it gives us a reasonable place to start. We hypothesized that women at greater risk of an adverse outcome might be more likely to benefit from the physician intervention. A bit to our surprise, we found the opposite result: lower-risk mothers appeared to experience larger decreases in mortality relative to higher-risk mothers. We speculate that this might be because the risks associated with these characteristics are harder to ameliorate in a primary health care setting. The marginal gains attached to the kind of 'low-tech' improvements in quality care that a physician might provide in a primary health setting might be larger for women without more complex health conditions.

In Columns 7-8 we test whether there were differential effects for women who were offered a conditional incentive compared to women who were not. We have previously shown that the incentive had strong effects on health care uptake (Okeke and Abubakar, 2020). Was there an interaction between the two treatments? The results are in Columns 7 and 8. We find similar effects in both samples. Finally, in Columns 9-10, we explore whether physician skill and medical technology might be complementary. Technology might be a multiplier in the sense that physicians might be able to do more when they have access to medical technology. To test this, we defined an index of facility capabilities. We assigned each health center a score based on the proportion of affirmative responses to the following questions: (a) were they connected to the power grid; did they have: (b) running water, (c) a functional backup generator, (d) at least 75% of a specified list of essential drugs, (e) at least 75% of a list of basic clinical equipment; and capacity to carry out: (f) hemoglobin tests, (g) urinalysis tests (h) blood transfusions, (i) neonatal resuscitation, and (j) caesarean sections. We compared outcomes for clusters with a health center in the top quartile of the capabilities index vs. the bottom three quartiles. The point estimates in both samples look quite similar. One implication of this result is that the primary constraint to improving outcomes may lie with human factors—in this case, the clinical ability of providers—rather than with technological limitations.[35]

In general, the effects of the physician treatment do not appear to be very different across different sub-groups. A joint overall test fails to reject the null of homogenous treatment effects.

4.7 Threats to validity

There are some potential threats to validity that are important to examine and, to the extent possible, rule out. One such threat is crossover between experimental arms. The main concern in this context would be crossover from control and MLP clusters to physician clusters because women learned that there was a physician available there. If present, crossover—from the other clusters to physician clusters—would imply that we are underestimating the treatment effect. We attempted to address the potential for crossover at the design stage by assigning treatment at the level of the health service area and by distributing participating sites across states so that clusters would not be particularly close together—as we noted previously, a local government area has, on average, two clusters—as a result the mean (median) distance from households

[35] See Das and Hammer (2014) for a discussion of the role of effort by health care providers as a determinant of outcomes.

in our sample to the nearest participating health center outside of the service area is 14 (12) kilometers. This limits the likelihood of crossover. This is borne out empirically by the data: first, we do not see any increase in physician-provided care in control and MLP clusters (see Figure 9), suggesting that there was no meaningful crossover, and second, in the follow-up survey where we asked women where they sought medical care—the options differentiated between the health center serving the cluster and any other health center—less than 2 percent of women overall reported giving birth in a health center other than the one in the cluster (1.6% in the control arm, 1.9% in MLP clusters, and 1.8% in physician clusters; p = 0.72).[36] We can therefore rule out crossover as a threat.

Another potential threat to validity is differential levels of monitoring. During the course of the intervention, we made repeated visits to the participating health center. As noted in Section 3, during these visits we collected various kinds of data, including taking provider roll calls. It is not a stretch to imagine that health centers and health providers may have seen these as monitoring visits and responded by becoming more diligent. If health centers in physician clusters received more visits, this might contribute to the observed treatment effect, implying that we would be overestimating the true effect of the physician treatment. Since all participating health centers received baseline and endline visits, the only remaining place for differentiation would be in the number of unannounced audit visits. Each time project staff visited the health center, for any reason, they were required to complete an audit form, so we know exactly how many times each health center was visited. In Figure A.10 we graph the mean number of audit visits by treatment arm. The data show that health centers in control clusters received about one visit less, on average, than health centers in the other arms, but there was no difference in the average number of visits to health centers in MLP and physician clusters. This suggests that differential monitoring is not a viable explanation for our results.

A third possible threat to validity might be differential provision of labor or capital inputs (other than the treatment) during the intervention period. If health centers in control or MLP clusters, for example, received additional resources, e.g., additional staff or improvements in infrastructure, and these contribute to improving patient outcomes, then we would be underestimating the treatment effect. Alternatively, if health centers in physician clusters were more likely to receive additional resources, then we might be overestimating the treatment effect. To examine this, we turn to the endline health facility survey where we collected information about whether any new workers had been posted to the health center since the baseline (this excludes the deployed provider): we define a variable equal to one if a health center received any additional staff. We also collected information about the condition of the building and other infrastructure such as tables and chairs, beds, screens, etcetera (as observed by research staff). Infrastructure upgrades or additional capital expenditure would potentially show up here. To assess condition, we used a four-point scale ranging from one (poor: "needing major rehabilitation") to four (excellent: "new or like new"). We graph the means of these variables and their 95% confidence intervals in Figure A.11. On all of these measures, we find no differences between arms, suggesting that we can cautiously rule out

[36] The proportion of women that reported visiting another health center for prenatal care was higher (6.7%). This is because smaller health centers known as health posts, which are sometimes closer, provide prenatal care but not delivery care. However, again, this percentage was not significantly different between arms (6.9% in the control arm, 7.6% in MLP clusters, and 5.7% in physician clusters; p = 0.25).

differential provision of human or capital resources.

4.8 Causal Chain

Our analysis has shown that easing physician supply constraints resulted in significant improvements in health outcomes. We have subjected the data to various tests to reassure ourselves that the results are real. Having shown this, next we turn our minds to the question of how these improvements in health were achieved. Our working hypothesis is that physicians are more skilled, and by substituting for less skilled health care providers, raised the level of quality, which in turn led to improvement in outcomes. In this section we systematically examine each link in this chain. Does the empirical evidence support this causal narrative? We begin with the first link: were the newly arrived physicians demonstrably more skilled, at diagnosing and managing medical conditions, than the existing mid-level providers whom they substituted for? Later in the section, we will also examine other complementary explanations.

4.8.1 Were the newly arrived physicians more skilled?

This is a particularly interesting question because the physicians that were sent to these communities were relatively young and inexperienced; the mean (and median) number of years of experience at the start of the intervention was two. In contrast, the average existing mid-level provider in our sample had a decade of experience working with their current medical qualification (mean/median of 11/10 years). One can think of a health care provider's level of clinical skill as a function of their intrinsic ability (raw intelligence), medical training, and clinical experience (on-the-job learning). While the physicians had more extensive training, they had less clinical experience, whereas the existing mid-level providers had less extensive training but many more years of on-the-job learning. This makes it an interesting exercise to compare their levels of clinical skill. We collected quite extensive data that allows us to do just this. As discussed in Section 3, we assessed health care providers—new and existing—on three domains: general medical knowledge, emergency obstetric case management (which is particularly germane given the outcome of interest), and outpatient primary care. For ease of interpretation, we report performance on each domain as a percentage score (out of 100).

We present these comparisons visually. In Figure 11, we plot kernel densities of the scores on each of these domains. We compare the densities for physicians, existing mid-level providers, and new mid-level providers. The latter had similar levels of experience to the physicians. The results are unequivocal: physicians outperformed mid-level providers in every area, and by a considerable margin.[37] This is similar to findings by others (Lohela et al., 2016). The distributions are quite distinct. Not only did the average physician demonstrate much higher levels of proficiency than the average mid-level provider, but even

[37] One concern, particularly with the case study and vignette results, is that they may be skewed by measures of process—history-taking and examinations—on which physicians are likely to perform much better. One way to address this is by focusing on the bottom line: rates of correct diagnosis and treatment. We examine this in Figure A.9 where we compare rates of correct diagnosis and treatment for the case of tuberculosis (as presented using the vignette). The case of pediatric malaria does not have sufficient discriminatory power as nearly all providers were able to correctly identify it. We see that physicians were about 18 percentage points more likely to make the correct diagnosis and 33 percentage points more likely to provide the correct treatment. If we include referrals as correct treatment the gap, compared to existing providers, narrows to a still large 17 percentage points.

low-performing physicians out-performed the majority of mid-level providers. For example, physicians at the 20th percentile—averaging over all of the scores—performed as well as mid-level providers at the 80th percentile. Having said that, it is worth pointing out that even though mid-level providers performed much worse on average, some mid-level providers demonstrated quite high levels of ability (the distributions are quite broad).[38] We can also see from the figures that experience does not seem to make a big difference; existing and new mid-level providers, despite the disparity in years of experiences, performed at similar levels.[39] In all cases, a Kolmogorov-Smirnov test cannot reject the null that the two distributions are the same. We conclude, based on these results, that physicians demonstrated considerably higher levels of clinical ability than mid-level health providers, thus establishing the first key link in the chain. Next, we examine the second link: did this raise the level of quality?

4.8.2 Did physicians deliver higher quality medical care?

While we may have established that physicians demonstrated higher levels of ability, this is necessary but not sufficient. There is a large literature showing that care in practice often lags behind what health care providers know to do. For an excellent review see Das and Hammer (2014).[40] It is therefore pertinent to examine whether patients in physician clusters received measurably better care. For this, we rely on our direct observations of care provided by physicians and mid-level providers. While what constitutes appropriate care will depend to some extent on the peculiarities of each case, it is possible to make broad comparisons. We use general metrics such as: (i) were physicians more likely to adhere to recommended history-taking protocols for common presentations such as fever. (ii) Were they more likely to perform a physical examination of the patient? We interpret this broadly. It includes checking the patient's temperature, measuring blood pressure, checking pulse rate, and checking for signs of anemia or dehydration. (iii) Were they more likely to make a diagnosis? To provide the right treatment, a provider must have, at least, a working diagnosis. We also use the length of the consultation as a marker of quality (Das and Hammer, 2014; Irving et al., 2017).[41] We use the natural log as the dependent variable.

Additionally, we look at multiple measures of provider-patient communication such as whether they explained a patient's diagnosis in common language, whether they explained the treatment being provided, and whether they gave any health education related to the diagnosis. We combine these into a single index by taking an average. Finally, we examine two 'bad' practices: propensity to prescribe injections (could be intramuscular or intravenous, vaccinations are excluded), and propensity to prescribe antibiotics. Parenteral drug administration should only be used when strictly necessary, but in developing country settings, it is not uncommon for health care providers to prescribe injections to signal that 'serious' treatment is being provided. It goes without saying that this is not good practice. Low quality health care providers tend to be more likely to indulge in this behavior. The over-use of antibiotics is also

[38]In Figure A.8 we examine heterogeneity by the type of mid-level provider.

[39]The relationship between years of experience and outcomes is fascinating. The common wisdom is that "practice makes perfect" but many studies find the opposite (Choudhry et al., 2005). There is some dispute about this and the answer remains unclear (Epstein et al., 2013; Goodwin et al., 2018; Tsugawa et al., 2017). Traoré et al. (2014) finds that experience is positively correlated with competence but Huchon et al. (2014), in the same setting, finds no correlation.

[40]For recent contributions see Mohanan et al. (2015), Das et al. (2016), and Okeke (2020).

[41]This is measured from the time the patient walked into the consulting room to when they exited.

a well-known problem (Chokshi et al., 2019; Ayukekbong et al., 2017). Prescribing antibiotics when not indicated is bad clinical practice and contributes to the problem of antibiotic resistance (Yip et al., 2014). For these last two indicators, lower rates are better.

To examine whether physicians provided better care, we estimate the following model using the directly observed outpatient care data. Y denotes a given quality metric for outpatient i seen in the health center serving cluster k:

$$Y_{ik} = \alpha + \beta_{1p} \sum_{p} ProviderType + \gamma X_{ik} + \theta_s + \epsilon_{ik} \tag{3}$$

$ProviderType$ indicates whether the health care provider was a physician, a new mid-level provider, or an existing mid-level provider. The latter are the omitted group. We control for a range of observable patient characteristics including age, sex, severity of illness, self-rated health, whether the patient presented with a fever, whether the visit was pregnancy-related (i.e., the patient was pregnant), the consultation order, whether the consultation was interrupted (e.g., because the patient was asked to go take a test and return with the result, or because the provider was called to attend to something else), phone ownership, and mode of transportation to the health center. The standard errors are adjusted for clustering at the health center level. We present the results in Table 8.

The results show that along nearly all dimensions, physicians provided a significant upgrade. Adherence to fever protocol was 15 percentage points higher, physicians were 8 percentage points more likely to have performed a physical examination, and 25 percentage points more likely to have made a diagnosis. Their 'soft' skills were also better: they scored about 5 percentage points higher on the communication index. As one might expect given these results, their consultations took longer (about 28% longer compared to existing mid-level providers). In Columns 6 and 7, where we examine indicators associated with bad clinical practice, they also performed better. Physicians were 8 and 11 percentage points less likely, respectively, to prescribe injections and antibiotics. We know that health care providers are likely to adjust their behavior in response to being observed (Leonard and Masatu, 2010; Okeke, 2020). If physicians responded more strongly to being observed, then these estimates represent an upper bound. We have some data on consultation times for unobserved consultations (patient entry/exit times was recorded by a research assistant stationed outside where they could observe the entrance to the consulting room). We can contrast differences between physicians and mid-level providers when they were observed vs. when they were not (see Table A.6). Indeed, consultations took longer when an observer was present, but the large difference between physicians and mid-level providers remains present even when there was no observer in the room.

Next we look specifically at obstetric care. Is there evidence that the quality of delivery care improved? This is particularly relevant because we know from the medical literature that intrapartum (delivery-related) events are responsible for about 1 in 3 newborn deaths (Liu et al., 2015). We rely on women's recall of birth events because it was not feasible for us to observe deliveries, and because health worker documentation in this setting is notoriously poor so we could not rely on health records (Bhattacharya et al., 2019). Depending on women's recall has limitations, but we are optimistic that the data can tell us something useful (especially since any reporting biases should be similar across the treatment arms). In designing the questionnaire, we relied on evidence from validation studies to guide which outcomes to include (and how to ask about them). Our use of a very short recall period—about 3 months on average—was also deliberate, based on evidence showing that recall was much more accurate when women were questioned within a few months after delivery (Bat-Erdene et al., 2013; Bhattacharya et al., 2019).

The key quality metric we will focus on is whether the woman was administered a uterotonic immediately after birth. Uterotonics are drugs that cause the uterus to contract. Uterine atony, or failure to contract, is a primary cause of severe postpartum bleeding, itself a leading cause of death (Say et al., 2014; Ayadi et al., 2013). Current guidelines recommend that a uterotonic agent (usually oxytocin) be administered prophylactically immediately after the birth of the child (Gallos et al., 2018; World Health Organization, 2018). It is usually administered as an intramuscular injection, though it can also be administered intravenously. In addition to its substantive importance, prior validation studies have shown that data on this metric can be reliably collected through population surveys (Blanc et al., 2016; Bhattacharya et al., 2019). In our endline survey, to ascertain whether women were administered a uterotonic, we asked if they received an intramuscular (or intravenous) injection immediately after giving birth.[42] We expect physicians to be more aware of the clinical benefit of administering a uterotonic agent and, therefore, to be more likely to prescribe its use. A supplementary metric that we also examine is whether a woman received cord traction to deliver the placenta. This is a component of a process known as "Active Management of the Third Stage of Labor" or AMTSL, which is recommended by the World Health Organization for all births (World Health Organization, 2017).

We present the intent-to-treat results in Table 9. For each outcome we report two specifications: one adjusting only for stratification and the other including additional controls. Consistent with improvements in delivery care, mothers in physician clusters were 4 percentage points more likely to have received a uterotonic agent immediately after birth (an 11 percent increase relative to the control group), and 5 percentage points more likely to have received cord traction (a 13 percent increase relative to the control group). These results indicate that there were substantive improvements in the quality of intrapartum care that help to explain the observed effect on health outcomes. They tell the same consistent story as the directly observed outpatient data, and together, present strong evidence that substituting physicians for mid-level providers led to improvements in the quality of patient care.

4.8.3 Clinic manager evaluations

To shed further light on how physicians may have impacted outcomes, we turn to a unique source of data: confidential evaluations of the new health care providers carried out by the health center managers (the in-charge). These evaluations were conducted towards the end of the posted providers' tenure. The health center managers were uniquely qualified to make these assessments for two reasons: first, they were health care providers themselves and so could make reasonably informed assessments of another health care provider (most were senior health care providers who had worked in multiple health facilities

[42] Misoprostol, which is sometimes used as a uterotonic, is administered orally, so we also asked if they had been given a drug to swallow or hold in their mouths immediately after birth. About 5% of women reported oral administration.

over their careers and would be able to make comparisons to other health providers that they had worked with before); second, they also had a front row seat, being able to observe these providers (and the patients they treated) up close, every day. These data provide valuable additional insight and add further nuance to the survey and observational data.

We asked the in-charges to rate the new provider's level of clinical knowledge, their level of skill in performing procedures, and their rapport with patients, along a scale from 0 (the lowest possible score) to 10 (the highest possible score). The phrasing of the question was: "On a scale of 0-10 where 10 is the highest possible score and 0 is the lowest, how would you rate [NAME] on [METRIC]? Separately we asked clinic managers to assess, using a Likert-like scale, what difference, if any, the provider had on service delivery in the health center. They could select one of the following options: strongly positive, fairly positive, neither positive nor negative, fairly negative, or strongly negative. We present the results in Figure 12.

The in-charges that worked with a physician were more likely to assign higher scores. On average, physicians were rated about 1 point higher (on a 10-point scale) than mid-level providers on the attributes of clinical knowledge and skill ($p < 0.001$). They were also rated about 0.8 points higher in terms of their rapport with patients ($p < 0.001$), which is consistent with our own independent observations of provider-patient interactions. Overall 69 percent of providers were rated as having had a strong positive impact, but there was a 15 percentage point swing between physicians and mid-level providers (77 vs. 62 percent: $p = 0.009$). Of course, these are subjective ratings and so it is possible that in-charges merely assigned physicians high scores on every metric simply because they were physicians, in which case these differences might not be very meaningful. We cannot rule this out, though the distributions show meaningful variation in scores. On clinical knowledge, for example, three physicians were rated a six or less.

To examine this in more depth, we collected additional information on the specific ways in which the health care provider might have impacted service delivery. For example, we asked the in-charge if the new provider introduced any innovations. If they responded in the affirmative, they were asked to give specific examples (the responses were open-ended). There are only 44 total responses so we report the full set of responses in Appendix Table A.7 with only minor edits for readability. The data show that, in health centers that received a physician, 43 percent of clinic managers reported that they introduced innovations, compared to 32 percent in mid-level provider health centers. Additionally, when we examine the specific kinds of innovations, there are important qualitative differences: for mid-level providers, the innovations were more general, e.g., giving advice about cleanliness of the facility or providing more health talks, whereas for physicians the innovations introduced were more pointed, and specific to patient care, e.g., better wound care and treatment of patients with high blood pressure.

When you put all of the data together, they tell a very compelling story: the arrival of physicians led to substantive measurable changes that help to explain why health outcomes improved.

4.9 Complementary mechanisms

4.9.1 Behavioral responses

Up until now, we have not examined whether there was a behavioral response to the intervention. One can imagine that individuals might change their behavior in response to the intervention in ways that might influence its effects on outcomes. We know that individuals care about, and respond to changes in quality of medical care (Leonard, 2007; Santos et al., 2017; Gaynor et al., 2016). Quality is, however, difficult to assess (Arrow, 1963). Individuals may therefore rely on observable proxies such as medical qualifications or age/experience. In low-information settings like this, being a physician may be taken as a credible signal of quality.[43] This suggests that individuals might respond to the presence of a physician in the community health center by switching from other sources of care. Of particular interest, given low levels of utilization of formal health care, is whether this crowds out informal provision of care. Substitution on the formal/informal margin is not only interesting but important, because if more women were drawn into the health care system in response to the physician treatment, this might help to amplify the intent-to-treat effect (because more women were effectively treated). We might also see substitution from other formal sources of care because of convenience: women who might otherwise have travelled to a general hospital because they wanted to have a physician present might now, instead, choose to use the primary health center.

We examine demand-side responses in Table 10. We examine prenatal care in Panel A, and place of birth in Panel B. To test for informal/formal substitution, we define indicators for a prenatal care in a health facility and for home birth (Column 1), and to examine substitution between formal care settings, we define a set of indicators for prenatal and delivery care in: (i) a public hospital, (ii) the primary health center, (iii) other public health facility, including another primary health center, (iv) a private hospital or clinic, and (v) other location (this includes churches and maternity homes).[44] We find a small decrease in the probability of home births in physician-assigned clusters—about 8 percent relative to the control group—and a compensating increase in births in the health center. But we also find a similar pattern in clusters assigned a mid-level provider, suggesting that this cannot explain differences in mortality.[45] We find no evidence of quantitatively important substitution across formal health care settings: all of the point estimates are uniformly close to zero. We also find little evidence of important changes in patterns of demand for prenatal care. One potential explanation is information frictions. At endline, over half of women in physician clusters were not aware that there was a physician available in the health center. Even among women who used the health center, one in three did not know. Over a longer period as information made its way through networks, one might expect to see larger behavioral effects.

4.9.2 Intensity of treatment

Another possibility is that physicians just have a more resource-intensive practice style. In this state of the world, it is not necessarily that physicians provide better care, it is that they provide more care when faced by the same patient. To explore this, we examine cost per case (a case here denotes a delivery) and length of stay (the number of nights spent in the facility for the delivery before the woman was allowed to

[43] This is one reason why informal health providers in the private sector are often anxious to describe themselves as 'doctors'.

[44] Some Pentecostal Christian churches encourage their congregants to give birth in the church so that they can receive prayer during this difficult time. Whether or not this leads to better outcomes is unclear.

[45] We speculate that this might be because an additional worker gives health centers more flexibility in providing service coverage in the health center, i.e., a volume effect. More predictable service coverage may lead to an increase in utilization (Chicoine and Guzman, 2017).

go home). Both of these measures capture the intensity of treatment. In the endline survey, we collected data on payments made for the delivery. Typical of expenditure data, the distribution is right-skewed with a mass at zero (26 percent of women reported zero costs). To analyze the effects on the cost of delivery, we used a two-part model: first modeling the probability of positive expenditure using a probit model, and then modeling costs conditional on some expenditure using GLM with a gamma distribution and log link (Mihaylova et al., 2011).[46] We report average marginal effects. For length of stay, we asked women how many nights they spent in the delivery facility (we excluded home births). It turns out that more than 60% of women were discharged home on the same day as the delivery, so we define an indicator for at least one night in the facility.[47] The results are shown in Table A.8. We do not find strong evidence of differences in costs or length of stay between physician clusters and other clusters, suggesting that, at least in this context, physicians did not merely provide more intensive care.

4.10 What are the returns to physician human capital?

So far our analysis of the effects of the policy experiment has focused on intent-to-treat effects but policy makers might be interested, not in the ITT effect, but in the average treatment effect, i.e., the average difference in outcomes between women who received medical care from a physician during pregnancy and childbirth, and women who received care from a mid-level provider. In other words, the health returns to a physician's additional human capital. This is difficult to estimate in practice in most settings because women whose care is provided by a physician are going to be very different from women whose care is provided by mid-level providers. For example, they might be more affluent women who have the resources to seek care in a private or government hospitals. The main issue for identification is that some of these differences will be difficult for an econometrician to observe and control for, even with very rich data.

Making such a comparison in this context, however, might be more plausible given that outside of the program physician-provided care was virtually non-existent. In effect, the probability that a woman's care was provided by a physician was almost entirely determined by whether she was lucky to live in a cluster that was randomly selected to receive a physician. Nevertheless, selection might still occur as a result of behavioral responses—individuals might alter their health-seeking behavior in response to the presence of a physician. The key issue is unmeasured heterogeneity that could be correlated with health outcomes. Table 10, however, shows little evidence of differential changes in health-seeking behavior—at least in physician relative to MLP clusters—likely because many women were unaware that a physician was now available in the health center. If we assume that there was no selection, an admittedly strong assumption that we will relax later, we can recover the average treatment effect using OLS.

We estimate a linear probability model where we define two new independent variables: one indicating whether a woman used formal health care services during her pregnancy and delivery, and an interaction between this dummy and another indicating whether she received care from a physician during pregnancy or childbirth. We rely, primarily, on information from women's health cards where physician-provided care

[46] We used the twopm stata module written by Belotti et al. (2015).

[47] There are important differences by facility type: in general hospitals women spend 2.4 nights, on average, in the facility; in private hospitals, 3.3 nights; in maternity homes, 2.5 nights, but in primary health centers, 0.4 nights on average.

is better measured (results based on the endline survey are reported in the Appendix). We estimate the following model:

$$Mortality_{iks} = \alpha + \beta_1 Usedcare_{iks} + \beta_2 Usedcare * Physician_{iks} + \gamma X_{iks} + \theta_s + \epsilon_{iks} \qquad (4)$$

The subscript iks identifies child i in cluster k in strata s. We include the same set of mother and child-level controls used previously. In this specification, β_1 measures the effect of using formal health care, when it is provided by a mid-level provider, compared to informal care. β_2 captures the incremental effect of medical care provided by a physician. Validity of β_2 requires that, conditional on using formal care, women whose care was, at least in part, provided by a physician are similar to women whose care was not. One way to examine the validity of this assumption is by comparing the average characteristics of women who saw a physician vs. a mid-level provider in the sample. We do this in Table A.9. On average, we see that they look quite similar; in fact we cannot reject the null that there is no selection.

Before presenting the results, it is worth discussing how $Usedcare_{iks}$ is defined. During pregnancy, a woman can use prenatal care or not; at the time of birth, she can also choose to go to a health care facility or not. There is some correlation between the two choices, for example 51 percent of women who attended prenatal care had a facility birth, compared to 5 percent of women that did not attend prenatal care, but there are a large number of women who use prenatal care but not delivery care (49 percent).[48] It is appealing to define formal care as using both prenatal and delivery care, but this would leave out the large number of women who used only prenatal care.[49] Prenatal care alone, even without facility care at birth, may offer health benefits. To balance these two considerations, we define formal care as having attended at least three prenatal visits or given birth in a health care facility. We assume that to receive the benefits of prenatal care, women must consume some minimum number. Four is the usual recommended minimum (World Health Organization et al., 2016), but we relax this and allow for at least three visits.

The results are in Table 11a. We find that women who saw a physician during pregnancy or childbirth experienced an incremental decrease of 1.2 percentage points in early newborn mortality relative to women who saw a mid-level provider during the pregnancy and childbirth. At the bottom of the table, we report the full physician effect $(\beta_1 + \beta_2)$ and test whether it is different from zero. Not surprisingly, given potential measurement error in the data, the estimated coefficients using information from the survey are smaller and biased towards zero (see Table A.12a). These estimates provide a useful benchmark but, as we have noted, rely on strong assumptions. At the cost of precision, we can estimate a local average treatment effect (or LATE) under weaker identification assumptions (Angrist and Pischke, 2009). We discuss this next.

There are two potentially endogenous variables in Equation (4), but the design of the policy experiment provides us with two exogenous instruments: the randomized offer of a conditional cash incentive and the randomized assignment of a physician. The former perturbs demand while the latter does the same for supply. We can therefore instrument for the second and third terms in Equation 4 using the randomized

[48]The reverse is not true. Only 2 percent of women who had a facility birth did not use any prenatal care.

[49]If we had the sample size it would have been interesting to disaggregate this further by comparing health care during both pregnancy and delivery to care during either pregnancy or delivery to no care.

offer of the conditional incentive and the interaction between the randomized offer and the dummy for assignment to the physician treatment arm respectively. They are, by design, exogenous. The interpretation of β_2 is as the effect of medical care for women who were induced to use care by the conditional incentive *and* who were able to see a physician because the health center in her service area was randomly assigned one.

A required assumption is that the conditional incentive did not have differential effects in the physician arm. If, for example, it had stronger effects on uptake in physician clusters, then the composition of users in physician clusters might be different. Table A.10 shows that the randomized incentive offer generated similar level effects in each arm. In Table A.11, we test directly for compositional effects by comparing the mean characteristics of users in each arm. Users in each arm consist of always takers and compliers—using the language of Angrist and Pischke (2009). Differences in average characteristics would indicate differences in the characteristics of compliers since, by definition, the fraction of always takers is the same in each arm. We see that users look quite similar in each arm. These results give us confidence that the randomized incentives did not generate differential effects.

The IV results are in Panel B of Table 11b. As expected, we pay a price in terms of the standard errors, but the results indicate that physician-provided medical care during pregnancy or childbirth, relative to care provided by a mid-level health care worker, reduced 1-day and 7-day infant mortality by an additional 2.8 and 3.6 percentage points respectively. The IV estimates are 2-3 times larger than the OLS estimates, indicative of negative selection into physician-provided care. It is also worth noting that the incremental effect when a physician provides care, is nearly as large as the main effect of formal care when it is provided by mid-level providers.

These results help to shed some light on a longstanding debate about whether shifting births into health facilities will, by itself, lead to reductions in high rates of child mortality in developing countries (Godlonton and Okeke, 2016; Das et al., 2018; Kruk et al., 2016a). The general direction of policy has been to push births into, what are often poorly resourced, health facilities (Kruk et al., 2016b). Several empirical papers have, however, argued that without making significant investments in improving quality, such policies are unlikely to make significant in-roads into infant mortality (Godlonton and Okeke, 2016; Okeke and Chari, 2014; Powell-Jackson et al., 2015). These results provide some support for this argument. They are consistent with findings by Godlonton and Okeke (2016) who examined the effect of a ban on informal birth attendants in Malawi, that had the same effect of shifting demand for formal health care services, and found reductions in newborn infant deaths only for women that resided close to a high-quality health care facility (based on a composite quality index).

We caution, however, that this does not mean that health care provided by mid-level providers has no value in terms of infant health. Okeke and Abubakar (2020) show that using medical care, even when it is provided by mid-level providers, reduces in utero mortality. This is plausibly explained by the fact that in utero deaths are largely determined by care during pregnancy. In general, prenatal care is not very complex care and can be adequately provided by less highly skilled health care providers, particularly with the increasing use of standardized protocols (McNabb et al., 2015). Early newborn mortality, on the other hand, as we have noted, is more closely linked to intrapartum events, which are less predictable and likely require higher levels of medical provider ability/training in order to navigate successfully.

4.11 Cost-benefit calculation

In thinking about the value of the physician program, it is important to realize that the mortality benefits, which we have shown, accrued not only to children in our sample, but potentially to all children born in physician clusters during the intervention period whose mothers used medical care. About 28 deliveries per month, on average, took place in health centers in our sample (data is from health center records at endline). This means that there were about 16,492 births, in total, in physician clusters over the 589 months that physicians were present. Assuming that 35% of these births received care from the physician during the pregnancy or delivery—the same proportion as in the sample for births in a health center that had a physician—this would imply that 208 newborn lives were saved by the intervention in physician clusters. This is calculated as 16,492 $\times$ 0.35 $\times$ 0.036 (the last number is taken from Table 11b Column 4). If we use the smaller proportion of births attended by a physician (15%), this implies that about 89 newborn lives were saved. For our cost-benefit calculations, we take the mid-point between these two numbers, 149.

We can value these lives in terms of earnings to get a sense of the benefits generated by the physician intervention. For simplicity, we ignore any non-mortality health benefits or any potential effects on adult mortality. To our knowledge there are no credible estimates of lifetime earnings for Nigeria, but we can carry out some back-of-the-envelope calculations to derive estimates. Current average life expectancy at birth in Nigeria is 54 years (World Bank Data Bank, 2020). Assuming an individual starts working at age 25, they can expect to work for 29 years. Adjusting for unemployment spells and job changes—we assume a 75% full-time equivalent—this gives 21.75 years of full-time earning. We assume mean annual earnings of $2,000 (equivalent to the mean gross national income per capita in 2019). Putting these together, we estimate that an individual in Nigeria would earn, on average, about $43,500 over their lifetime. For comparison, average lifetime earnings in the US, over a 50-year working life, exceed $1 million dollars (Tamborini et al., 2015). Multiplying the value of lifetime earnings by the number of lives saved, suggests that the program generated about $6.5 million in value. This stream of income is many years in the future, so one might want the net present value. Applying an annual real discount rate of 4%, we calculate the net present value to be about $1.7 million.

How does this compare to the cost of the intervention? We start with the variable costs. The largest cost of the intervention was the physicians' salaries. Total salary costs were $122,826 ($208 per physician per month). Accommodation for the physicians was usually provided for free to the physicians by the local government, but valued at prevailing market rental rates, the estimated cost was $81,806.[50] We also include monitoring costs: we assume that it would be necessary to visit each health facility where a physician was posted, at least once every three months. The estimated total cost of monitoring is $10,907 ($18.50 per facility per month). Implementation costs, including planning and coordination meetings came to approximately $8,056 (the cost per facility was $139). These are fixed costs that would be incurred in setting up a similar program. Putting all these costs together, the total cost of the intervention comes to $223,595

[50]We assumed an average monthly rental rate of $140 for a 2-bedroom apartment.

(approximately \$1,500 per life saved). This implies that every \$1 invested in the program returned \$7.50 in benefits. Even under the most conservative scenario in which only 89 lives were saved, the program returned \$4.50 in benefits for every dollar spent.

Our calculations are based on actual paid salary costs. As we have noted, however, these salaries were less than what physicians would earn on the open market, outside of the program. In an alternative scenario in which physicians are being recruited for these positions, one can easily imagine that physicians would demand a significant wage premium to work in underserved areas (Grobler et al., 2015; Bärnighausen and Bloom, 2009). Monthly wages for a post-NYSC medical officer in Nigeria generally range from \$400-\$700. Let's say that in order to make the position attractive, physicians were paid \$1,000 a month, this would increase the total cost of the intervention to \$607,963. Even at this number, the program still would return \$2.80 in benefits for every dollar spent under realistic assumptions. The cost per life saved would increase to \$4,094, or about \$76 per life year. This would still make it remarkably cost-effective. For reference, oral rehydration therapy is estimated to cost about \$200 per quality-adjusted year (Horton et al., 2017).

5 Concluding remarks

This paper has presented results from a unique policy experiment in Nigeria in which physicians were randomly posted to medically underserved communities for one year. 180 primary health service areas, drawn from five different states, were selected to take part in the experiment. 60 of these were randomly selected to receive a physician. The physician was posted to the government primary health care center serving the area, joining the existing complement of health care workers working in the health center (only 2 out of the 180 health centers had a physician on staff at baseline). To identify the 'quality' effect of a physician, separate from the 'quantity' effect of receiving an additional health care worker, another 60 service areas were randomly selected to contemporaneously receive a mid-level health care provider. The mid-level provider had similar qualifications to existing health care workers in the health center. The remaining 60 service areas did not receive any additional health care workers and served as a control group.

Our results indicate that the physician program led to a significant reduction in infant mortality. In primary health services areas assigned a physician, 1-day and 7-day infant mortality reduced by between 6-8 deaths per thousand live births on an intent-to-treat basis. These results are striking given that the physicians were there for less than on year. One would expect the effects to be larger over a longer period. We found that the size of the mortality effect increased with longer duration of exposure. Infants with longer than median exposure—based on the number of pregnancy months exposed—experienced a decrease of between 9-12 deaths per thousand and 11-13 deaths per thousand in 1-day and 7-day infant mortality, respectively. In our analysis of mechanisms, we showed that physicians represented a significant upgrade in quality from mid-level providers, helping to explain why outcomes improved.

It is likely that there are other benefits that are not captured in our analysis. We did not examine maternal mortality, for example, because of sample size limitations. It is not unreasonable to think that women would also benefit from improved access to physicians. Similarly, we did not examine the effect on the outcomes of other children and adults who might also have received care from these health centers during the intervention period. Lastly, there might be skill diffusion from physicians to mid-level providers in the same health center.[51] Such effects would last long beyond the tenure of the physician. In that sense this paper only captures some of the potential benefits of easing physician supply constraints. We leave these issues for future work.

This paper highlights the dangers of task-shifting policy as a strategy for dealing with physician shortages. A better short run strategy might be to redistribute existing physicians (though this may not be possible everywhere). Getting physicians to permanently move to underserved areas, even with incentives, may not be practical, but providing coverage of high-need areas using a rotating corp of physicians on short-term (or even ultra short-term) postings may be an acceptable (and feasible) alternative. One could also imagine incorporating rural health postings into residency training programs.[52] This would involve requiring resident doctors to do short postings in rural health facilities as part of their curriculum. The recent coronavirus pandemic has also highlighted some other innovative strategies for dealing with shortages, such as using retired physicians or medical students in their final year of training. In the long run, however, addressing supply constraints is hard to do without expanding the production of physicians.

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. (2017). When should you adjust standard errors for clustering? Working Paper 24003, National Bureau of Economic Research.

Adhvaryu, A. and Nyshadham, A. (2015). Returns to treatment in the formal health care sector: Evidence from Tanzania. *American Economic Journal: Economic Policy*, 7(3):29–57.

Anderson, D. M., Brown, R., Charles, K. K., and Rees, D. I. (2020). Occupational licensing and maternal health: Evidence from early midwifery laws. *Journal of Political Economy.*, Forthcoming.

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

Arrow, K. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, 53(3):941–73.

Assocation of American Medical Colleges (AAMC) (2019). Results of the 2018 Medical School Enrollment Survey. *Assocation of American Medical Colleges, Washington DC.*

Assocation of American Medical Colleges (AAMC) (2020). The Complexities of Physician Supply and Demand: Projections from 2018-2033. *Association of American Medical Colleges Washington, D.C. Report prepared for the AAMC by IHS Markit Ltd.*

Astolfi, P. and Zonta, L. A. (1999). Risks of preterm delivery and association with maternal age, birth order, and fetal gender. *Human Reproduction*, 14(11):2891–2894.

[51] This is the subject of ongoing work.

[52] Some of this happens already with community medicine/public health residency training but could be extended to all trainees.

Australia, H. W. (2012). Health Workforce 2025 – Volumes 1 to 3. Technical report, Health Workforce Australia, Canberra.

Ayadi, A. M. E., Robinson, N., Geller, S., and Miller, S. (2013). Advances in the treatment of postpartum hemorrhage. *Expert Review of Obstetrics & Gynecology*, 8(6):525–537.

Ayukekbong, J. A., Ntemgwa, M., and Atabe, A. N. (2017). The threat of antimicrobial resistance in developing countries: causes and control strategies. *Antimicrobial Resistance & Infection Control*, 6(1):47.

Baicker, K. and Chandra, A. (2018). Challenges in Understanding Differences in Health Care Spending Between the United States and Other High-Income Countries. *JAMA*, 319(10):986–987.

Bärnighausen, T. and Bloom, D. E. (2009). Financial incentives for return of service in underserved areas: a systematic review. *BMC Health Services Research*, 9(1):86.

Bartel, A. P., Beaulieu, N. D., Phibbs, C. S., and Stone, P. W. (2014). Human capital and productivity in a team environment: Evidence from the healthcare sector. *American Economic Journal: Applied Economics*, 6(2):231–59.

Bat-Erdene, U., Metcalfe, A., McDonald, S. W., and Tough, S. C. (2013). Validation of canadian mothers'recall of events in labour and delivery with electronic health records. *BMC Pregnancy and Childbirth*, 13(1):S3.

Belotti, F., Deb, P., Manning, W. G., and Norton, E. C. (2015). twopm: Two-part models. *The Stata Journal*, 15(1):3–20.

Bhattacharya, A. A., Allen, E., Umar, N., Usman, A. U., Felix, H., Audu, A., Schellenberg, J. R., and Marchant, T. (2019). Monitoring childbirth care in primary health facilities: a validity study in Gombe State, northeastern Nigeria. *Journal of global health*, 9(2):020411–020411.

Bhutta, Z. A., Das, J. K., Bahl, R., Lawn, J. E., Salam, R. A., Paul, V. K., Sankar, M. J., Blencowe, H., Rizvi, A., Chou, V. B., and Walker, N. (2014). Can available interventions end preventable deaths in mothers, newborn babies, and stillbirths, and at what cost? *The Lancet*, 384(9940):347 – 370.

Blanc, A. K., Diaz, C., McCarthy, K. J., and Berdichevsky, K. (2016). Measuring progress in maternal and newborn health care in Mexico: validating indicators of health system contact and quality of care. *BMC Pregnancy and Childbirth*, 16(1):255.

Carrillo, B. and Feres, J. (2019). Provider supply, utilization, and infant health: evidence from a physician distribution policy. *American Economic Journal: Economic Policy*, 11(3):156–96.

Chandra, A. and Skinner, J. (2012). Technology growth and expenditure growth in health care. *Journal of Economic Literature*, 50(3):645–80.

Chicoine, L. E. and Guzman, J. C. (2017). Increasing rural health clinic utilization with SMS updates: Evidence from a randomized evaluation in Uganda. *World Development*, 99:419–430.

Chokshi, A., Sifri, Z., Cennimo, D., and Horng, H. (2019). Global contributors to antibiotic resistance. *Journal of global infectious diseases*, 11(1):36.

Choudhry, N. K., Fletcher, R. H., and Soumerai, S. B. (2005). Systematic review: The relationship between clinical experience and quality of health care. *Annals of Internal Medicine*, 142(4):260–273.

Cooper, R. A., Getzen, T. E., McKee, H. J., and Laud, P. (2002). Economic and demographic trends signal an impending physician shortage. *Health Affairs*, 21(1):140–154. PMID: 11900066.

Das, J. and Hammer, J. (2014). Quality of primary care in low-income countries: Facts and economics. *Annual Review of Economics*, 6(1):525–553.

Das, J., Holla, A., Mohpal, A., and Muralidharan, K. (2016). Quality and accountability in health care delivery: audit-study evidence from primary care in India. *American Economic Review*, 106(12):3765–3799.

Das, J., Woskie, L., Rajbhandari, R., Abbasi, K., and Jha, A. (2018). Rethinking assumptions about delivery of healthcare: implications for universal health coverage. *BMJ*, 361.

de Brey, C., Musu, L., McFarland, J., Wilkinson-Flicker, S., Diliberti, M., Zhang, A., Branstetter, C., and Wang, X. (2018). Indicator 1: Population distribution. In *Status and Trends in the Education of Racial and Ethnic Groups 2018*. National Center for Education Statistics.

Dieleman, J. L., Cao, J., Chapin, A., Chen, C., Li, Z., Liu, A., Horst, C., Kaldjian, A., Matyasz, T., Scott, K. W., Bui, A. L., Campbell, M., Duber, H. C., Dunn, A. C., Flaxman, A. D., Fitzmaurice, C., Naghavi, M., Sadat, N., Shieh, P., Squires, E., Yeung, K., and Murray, C. J. L. (2020). US Health Care Spending by Payer and Health Condition, 1996-2016. *JAMA*, 323(9):863–884.

Doyle, J. J., Ewer, S. M., and Wagner, T. H. (2010). Returns to physician human capital: Evidence from patients randomized to physician teams. *Journal of Health Economics*, 29(6):866 – 882.

Epstein, A. J., Srinivas, S. K., Nicholson, S., Herrin, J., and Asch, D. A. (2013). Association between physicians' experience after training and maternal obstetrical outcomes: cohort study. *BMJ*, 346.

European Commission (2012). Commission staff working document on an action plan for the EU health workforce. Technical report, European Commission, Brussels.

Evans, W. N. and Lien, D. S. (2005). The benefits of prenatal care: evidence from the PAT bus strike. *Journal of Econometrics*, 125(1-2):207–239.

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., Allen, H., Baicker, K., and Group, O. H. S. (2012). The Oregon health insurance experiment: evidence from the first year. *The Quarterly Journal of Economics*, 127(3):1057–1106.

Fontes, L. F. C., Conceição, O. C., and Jacinto, P. d. A. (2018). Evaluating the impact of physicians' provision on primary healthcare: Evidence from Brazil's More Doctors Program. *Health Economics*, 27(8):1284–1299.

Gabrysch, S., Nesbitt, R. C., Schoeps, A., Hurt, L., Soremekun, S., Edmond, K., Manu, A., Lohela, T. J., Danso, S., Tomlin, K., Kirkwood, B., and Campbell, O. M. R. (2019). Does facility birth reduce maternal and perinatal mortality in Brong Ahafo, Ghana? a secondary analysis using data on 119,244 pregnancies from two cluster-randomised controlled trials. *The Lancet Global Health*, 7(8):e1074–e1087.

Gajate-Garrido, G. (2013). The impact of adequate prenatal care on urban birth outcomes: An analysis in a developing country context. *Economic Development and Cultural Change*, 62(1):95–130.

Gallos, I. D., Papadopoulou, A., Man, R., Athanasopoulos, N., Tobias, A., Price, M. J., Williams, M. J., Diaz, V., Pasquale, J., Chamillard, M., et al. (2018). Uterotonic agents for preventing postpartum haemorrhage: a network meta-analysis. *Cochrane Database of Systematic Reviews*, (12).

Gaynor, M., Propper, C., and Seiler, S. (2016). Free to choose? reform, choice, and consideration sets in the English National Health Service. *American Economic Review*, 106(11):3521–57.

Godlonton, S. and Okeke, E. N. (2016). Does a ban on informal health providers save lives? evidence from Malawi. *Journal of Development Economics*, 118:112 – 132.

Goldstein, M., Graff Zivin, J., Habyarimana, J., Pop-Eleches, C., and Thirumurthy, H. (2013). The effect of health worker absence and health clinic protocol on health outcomes: the case of mother-to-child transmission of HIV in Kenya. *American Economic Journal: Applied Economics*, 5:58–85.

Gonzalez, F. and Kumar, S. (2018). Prenatal care and birthweight in Mexico. *Applied Economics*, 50(10):1156–1170.

Goodman, D. C. and Robertson, R. G. (2013). Accelerating physician workforce transformation through competitive graduate medical education funding. *Health Affairs*, 32(11):1887–1892.

Goodman-Bacon, A. (2018). Public insurance and mortality: evidence from Medicaid implementation. *Journal of Political Economy*, 126(1):216–262.

Goodwin, J. S., Salameh, H., Zhou, J., Singh, S., Kuo, Y.-F., and Nattinger, A. B. (2018). Association of Hospitalist Years of Experience With Mortality in the Hospitalized Medicare Population. *JAMA Internal Medicine*, 178(2):196–203.

Grobler, L., Marais, B. J., and Mabunda, S. (2015). Interventions for increasing the proportion of health professionals practising in rural and other underserved areas. *Cochrane database of systematic reviews*, (6).

Health Resources and Services Administration (2020). Designated Health Professional Shortage Areas Statistics. Third Quarter of Fiscal Year 2020 Designated HPSA Quarterly Summary. Technical report, Bureau of Health Workforce Health Resources and Services Administration (HRSA) U.S. Department of Health & Human Services. Available at `https://data.hrsa.gov/Default/GenerateHPSAQuarterlyReport`.

Healthy People 2020. Healthy people 2020: Access to health services. HealthyPeople.gov https://www.healthypeople.gov/2020/topics-objectives/topic/Access-to-Health-Services.

Horton, S., Gelband, H., Jamison, D., Levin, C., Nugent, R., and Watkins, D. (2017). Ranking 93 health interventions for low-and middle-income countries by cost-effectiveness. *PloS one*, 12(8):e0182951.

Huchon, C., Arsenault, C., Tourigny, C., Coulibaly, A., Traore, M., Dumont, A., and Fournier, P. (2014). Obstetric competence among referral healthcare providers in Mali. *International Journal of Gynecology & Obstetrics*, 126(1):56–59.

Irving, G., Neves, A. L., Dambha-Miller, H., Oishi, A., Tagashira, H., Verho, A., and Holden, J. (2017). International variations in primary care physician consultation time: a systematic review of 67 countries. *BMJ Open*, 7(10).

Ishikawa, T., Ohba, H., Yokooka, Y., Nakamura, K., and Ogasawara, K. (2013). Forecasting the absolute and relative shortage of physicians in japan using a system dynamics model approach. *Human resources for health*, 11(1):41.

Kaiser Family Foundation (2019). Total number of medical school graduates. Available at `https://www.kff.org/other/state-indicator/total-medical-school-graduates`.

Kleiner, M. M. (2016). Battling over jobs: occupational licensing in health care. *American Economic Review*, 106(5):165–70.

Kleiner, M. M., Marier, A., Park, K. W., and Wing, C. (2016). Relaxing occupational licensing requirements: Analyzing wages and prices for a medical service. *The Journal of Law and Economics*, 59(2):261–291.

Kolstad, J. T. and Kowalski, A. E. (2012). The impact of health care reform on hospital and preventive care: evidence from Massachusetts. *Journal of Public Economics*, 96(11-12):909–929.

Kotlikoff, L. J. and Hagist, C. (2005). Who's going broke? comparing growth in healthcare costs in ten OECD countries. *NBER Working Paper*, (w11833).

Kruk, M. E., Larson, E., and Twum-Danso, N. A. Y. (2016a). Time for a quality revolution in global health. *The Lancet Global Health*, 4(9):e594 – e596.

Kruk, M. E., Leslie, H. H., Verguet, S., Mbaruku, G. M., Adanu, R. M. K., and Langer, A. (2016b). Quality of basic maternal care functions in health facilities of five african countries: an analysis of national health system surveys. *The Lancet Global Health*, 4(11):e845 – e855.

Kupka, R., Kassaye, T., Saathoff, E., Hertzmark, E., Msamanga, G. I., and Fawzi, W. W. (2009). Predictors of stillbirth among HIV-infected Tanzanian women. *Acta obstetricia et gynecologica Scandinavica*, 88(5):584–592.

Laurant, M., van der Biezen, M., Wijers, N., Watananirun, K., Kontopantelis, E., and van Vught, A. J. (2018). Nurses as substitutes for doctors in primary care. *Cochrane Database of Systematic Reviews*, (7).

Lawn, J. E., Blencowe, H., Oza, S., You, D., Lee, A. C., Waiswa, P., Lalli, M., Bhutta, Z., Barros, A. J., Christian, P., et al. (2014). Every newborn: progress, priorities, and potential beyond survival. *The Lancet*, 384(9938):189–205.

Lawn, J. E., Cousens, S., and Zupan, J. (2005). 4 million neonatal deaths: When? where? why? *The Lancet*, 365(9462):891 – 900.

Leonard, K. L. (2007). Learning in health care: Evidence of learning about clinician quality in Tanzania. *Economic Development and Cultural Change*, 55(3):531–55.

Leonard, K. L. and Masatu, M. C. (2010). Using the Hawthorne effect to examine the gap between a doctors best possible practice and actual practice. *Journal of Development Economics*, 93(2):226–243.

Liu, J. X., Goryakin, Y., Maeda, A., Bruckner, T., and Scheffler, R. (2016). *Global Health Workforce Labor Market Projections for 2030.* The World Bank.

Liu, L., Oza, S., Hogan, D., Perin, J., Rudan, I., Lawn, J. E., Cousens, S., Mathers, C., and Black, R. E. (2015). Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an updated systematic analysis. *The Lancet*, 385(9966):430–440.

Lohela, T. J., Nesbitt, R. C., Manu, A., Vesel, L., Okyere, E., Kirkwood, B., and Gabrysch, S. (2016). Competence of health workers in emergency obstetric care: an assessment using clinical vignettes in Brong Ahafo region, Ghana. *BMJ Open*, 6(6).

Markowitz, S., Adams, E. K., Lewitt, M. J., and Dunlop, A. L. (2017). Competitive effects of scope of practice restrictions: Public health or public harm? *Journal of Health Economics*, 55:201 – 218.

McNabb, M., Chukwu, E., Ojo, O., Shekhar, N., Gill, C. J., Salami, H., and Jega, F. (2015). Assessment of the quality of antenatal care services provided by health workers using a mobile phone decision support application in northern Nigeria: a pre/post-intervention study. *PLOS one*, 10(5):e0123940.

McPake, B. and Mensah, K. (2008). Task shifting in health care in resource-poor countries. *The Lancet*, 372(9642):870–871.

Medical and Dental Council of Nigeria (2019). Status of Accredited Medical and Dental Schools in Nigeria. Available at `https://bit.ly/31vHJRL`.

Mihaylova, B., Briggs, A., O'Hagan, A., and Thompson, S. G. (2011). Review of statistical methods for analysing healthcare resources and costs. *Health Economics*, 20(8):897–916.

Miller, S. and Wherry, L. R. (2017). Health and access to care during the first 2 years of the ACA Medicaid expansions. *New England Journal of Medicine*, 376(10):947–956.

Mills, E. J., Kanters, S., Hagopian, A., Bansback, N., Nachega, J., Alberton, M., Au-Yeung, C. G., Mtambo, A., Bourgeault, I. L., Luboga, S., Hogg, R. S., and Ford, N. (2011). The financial cost of doctors emigrating from sub-Saharan Africa: human capital analysis. *BMJ*, 343.

Moberly, T. (2017). UK has fewer doctors per person than most other OECD countries. *BMJ*, 357.

Mohanan, M., Vera-Hernández, M., Das, V., Giardili, S., Goldhaber-Fiebert, J. D., Rabin, T. L., Raj, S. S., Schwartz, J. I., and Seth, A. (2015). The know-do gap in quality of health care for childhood diarrhea and pneumonia in rural India. *JAMA pediatrics*, 169(4):349–357.

Naeye, R. L., Burt, L. S., Wright, D. L., Blanc, W. A., and Tatter, D. (1971). Neonatal mortality, the male disadvantage. *Pediatrics*, 48(6):902–906.

National Bureau of Statistics (2018). Joint Admissions and Matriculation Board (JAMB) Candidates by state 2017-2018. Available from: `https://nigerianstat.gov.ng/download/917`.

National Population Commission and ICF International (2014). Nigeria demographic and health survey 2013. Technical report, NPC and ICF International, Abuja, Nigeria, and Rockville, Maryland, USA.

National Population Commission and ICF International (2019). Nigeria Demographic and Health Survey 2018. Technical report, NPC and ICF International, Abuja, Nigeria, and Rockville, Maryland, USA.

National Primary Health Care Development Agency (2014). Minimum standards for primary health care in Nigeria. Technical report, Department of Planning, Research and Statistics, NPHCDA Plot 681/682 Port-Harcourt Crescent, Off Gimbiya Street, Garki II, Abuja.

Nigerian National Bureau of Statistics (2016). LSMS-Integrated Surveys on Agriculture General Household Survey Panel 2015/2016. Technical report, Nigerian National Bureau of Statistics in Collaboration with the Federal Ministry of Agriculture and Rural Development and the World Bank.

Okeke, E. (2020). Working hard or hardly working: Health worker effort and health outcomes. *Economic Development and Cultural Change*.

Okeke, E., Glick, P., Chari, A., Abubakar, I., Pitchforth, E., Exley, J., Bashir, U., Gu, K., and Onwujekwe, O. (2016). The effect of increasing the supply of skilled health providers on pregnancy and birth outcomes: Evidence from the Midwives Service Scheme in Nigeria. *BMC Health Services Research*, 16(1):425.

Okeke, E. N. and Abubakar, I. S. (2020). Healthcare at the beginning of life and child survival: Evidence from a cash transfer experiment in Nigeria. *Journal of Development Economics*, 143:102426.

Okeke, E. N. and Chari, A. (2014). Can Institutional Deliveries Reduce Newborn Mortality? Evidence from Rwanda. *RAND Working Paper 1072*.

Okeke, E. N., Pitchforth, E., Exley, J., Glick, P., Abubakar, I. S., Chari, A. V., Bashir, U., Gu, K., and Onwujekwe, O. (2017). Going to scale: design and implementation challenges of a program to increase access to skilled birth attendants in Nigeria. *BMC health services research*, 17(1):356.

Okonofua, F., Lambo, E., Okeibunor, J., and Agholor, K. (2011). Advocacy for free maternal and child health care in Nigeria—results and outcomes. *Health Policy*, 99(2):131 – 138.

Ordinioha, B. and Onyenaporo, C. (2010). Experience with the use of community health extension workers in primary care, in a private rural health care institution in South-South Nigeria. *Annals of African Medicine*, 9(4):240–245.

Papanicolas, I., Woskie, L. R., and Jha, A. K. (2018). Health Care Spending in the United States and Other High-Income Countries. *JAMA*, 319(10):1024–1039.

Perloff, J., DesRoches, C. M., and Buerhaus, P. (2016). Comparing the cost of care provided to Medicare beneficiaries assigned to primary care nurse practitioners and physicians. *Health Services Research*, 51(4):1407–1423.

Petterson, S. M., Liaw, W. R., Phillips, R. L., Rabin, D. L., Meyers, D. S., and Bazemore, A. W. (2012). Projecting us primary care physician workforce needs: 2010-2025. *The Annals of Family Medicine*, 10(6):503–509.

Powell-Jackson, T., Mazumdar, S., and Mills, A. (2015). Financial incentives in health: New evidence from India's Janani Suraksha Yojana. *Journal of Health Economics*, 43:154–169.

Rama, A. (2020). National Health Expenditures, 2018: Spending Growth Remains Steady Even With Increases in Private Health Insurance and Medicare Spending. *American Medical Association: Policy Research Perspectives*.

Sankar, M., Natarajan, C., Das, R., Agarwal, R., Chandrasekaran, A., and Paul, V. (2016). When do newborns die? a systematic review of timing of overall and cause-specific neonatal deaths in developing countries. *Journal of Perinatology*, 36(1):S1–S11.

Santos, R., Gravelle, H., and Propper, C. (2017). Does quality affect patients' choice of doctor? evidence from England. *The Economic Journal*, 127(600):445–494.

Say, L., Chou, D., Gemmill, A., Tunçalp, Ö., Moller, A., Daniels, J., Gülmezoglu, A. M., Temmerman, M., and Alkema, L. (2014). Global causes of maternal death: a WHO systematic analysis. *The Lancet Global Health*, 2(6):e323–e333.

Scheffler, R. M., Mahoney, C. B., Fulton, B. D., Dal Poz, M. R., and Preker, A. S. (2009). Estimates of health care professional shortages in Sub-Saharan Africa by 2015: Critical shortages of health workers are projected in 2015 throughout sub-Saharan Africa, and the projected cost to eliminate them approaches $20 billion. *Health Affairs*, 28(Suppl1):w849–w862.

Shigeoka, H. (2014). The effect of patient cost sharing on utilization, health, and risk protection. *American Economic Review*, 104(7):2152–84.

Sommers, B. D., Baicker, K., and Epstein, A. M. (2012). Mortality and access to care among adults after state medicaid expansions. *New England Journal of Medicine*, 367(11):1025–1034.

Tamborini, C. R., Kim, C., and Sakamoto, A. (2015). Education and lifetime earnings in the United States. *Demography*, 52(4):1383–1407.

The Lancet (2018). GBD 2017: a fragile world. *The Lancet*, 392(10159):1683.

Traczynski, J. and Udalova, V. (2018). Nurse practitioner independence, health care utilization, and health outcomes. *Journal of Health Economics*, 58:90–109.

Traoré, M., Arsenault, C., Schoemaker-Marcotte, C., Coulibaly, A., Huchon, C., Dumont, A., and Fournier, P. (2014). Obstetric competence among primary healthcare workers in Mali. *International Journal of Gynecology & Obstetrics*, 126(1):50–55.

Tsugawa, Y., Newhouse, J. P., Zaslavsky, A. M., Blumenthal, D. M., and Jena, A. B. (2017). Physician age and outcomes in elderly patients in hospital in the US: observational study. *BMJ*, 357.

United Nations Children's Fund (2019). Levels and trends in child mortality. report 2019. Estimates developed by the UN inter-agency group for child mortality estimation.

Vogel, J. P., Torloni, M. R., Seuc, A., Betrán, A. P., Widmer, M., Souza, J. P., and Merialdi, M. (2013). Maternal and perinatal outcomes of twin pregnancy in 23 low- and middle-income countries. *PLoS ONE*, 8(8):e70549.

Watkins, W. J., Kotecha, S. J., and Kotecha, S. (2016). All-cause mortality of low birthweight infants in infancy, childhood, and adolescence: Population study of England and Wales. *PLOS Medicine*, 13(5):1–19.

World Bank (2020). *World Development Indicators, World Bank, Washington.*

World Bank Data Bank (2020). Mortality rate, neonatal (per 1,000 live births), 2017. Technical report, World Bank. Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation.

World Health Organization (2014). A universal truth: no health without a workforce. Technical report, World Health Organization.

World Health Organization (2016). Health workforce requirements for universal health coverage and the Sustainable Development Goals. *World Health Organization. Background paper No. 1 to the Global Strategy on Human Resources for Health.*

World Health Organization (2017). WHO recommendations on maternal health: guidelines approved by the WHO Guidelines Review Committee. Technical report, World Health Organization, Geneva.

World Health Organization (2018). WHO recommendations: uterotonics for the prevention of postpartum haemorrhage. Technical report, World Health Organization, Geneva. Licence: CC BY-NC-SA 3.0 IGO.

World Health Organization (2019a). Newborns: Reducing mortality. *WHO Fact Sheet. https://www.who.int/news-room/fact-sheets/detail/newborns-reducing-mortality. Accessed July 14, 2020.*

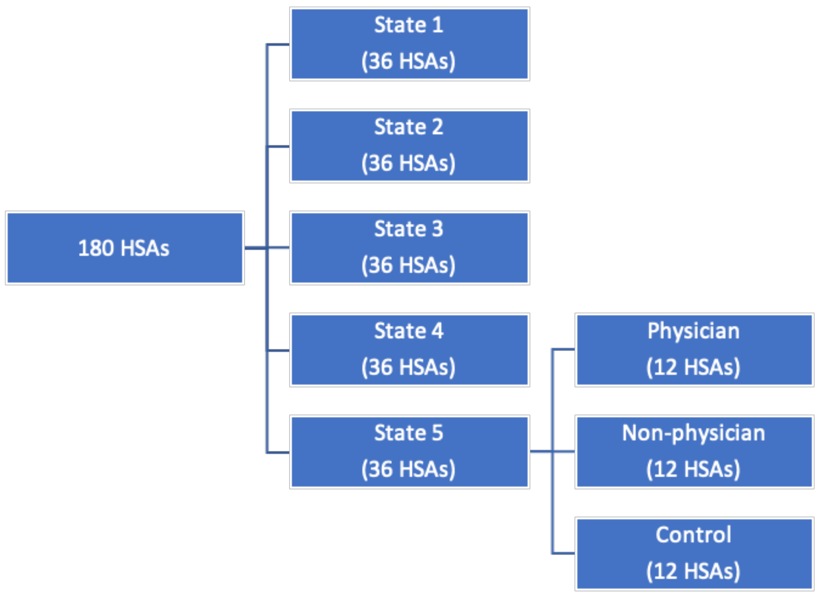
World Health Organization (2019b). Trends in maternal mortality 2000 to 2017: estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division. *World Health Organization.*

World Health Organization et al. (2016). *WHO recommendations on antenatal care for a positive pregnancy experience*. World Health Organization.

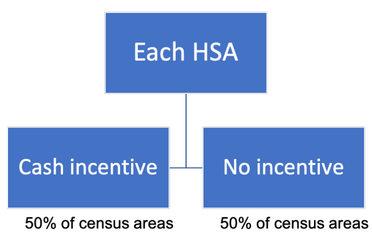
Yip, W., Powell-Jackson, T., Chen, W., Hu, M., Fe, E., Hu, M., Jian, W., Lu, M., Han, W., and Hsiao, W. C. (2014). Capitation combined with pay-for-performance improves antibiotic prescribing practices in rural China. *Health Affairs*, 33(3):502–510.

Young, A., Chaudhry, H. J., Pei, X., Arnhart, K., Dugan, M., and Snyder, G. B. (2019). FSMB census of licensed physicians in the United States, 2018. *Journal of Medical Regulation*, 105(2):7–23.

Figure 1: Experimental design



(a) Health worker intervention



(b) Conditional incentive

Figure 1 lays out the experimental design. A HSA denotes a Primary Health Service Area, which consists of communities served by a government primary health center. Figure 1a shows the design of the primary intervention: 180 HSAs were randomly assigned with equal probability to one of three experimental arms: Physician denotes clusters randomly assigned a new physician, MLP denotes clusters randomly assigned a new mid-level health care provider, and Control denotes clusters that received no additional workers. Health workers were posted to the primary health center serving the cluster. A secondary intervention was nested within this (see Figure 1b): in each participating HSA, 50% of census areas were randomly assigned to an intervention in which pregnant women were offered a payment of $14, to be made after the birth of their child, if the woman attended at least 3 prenatal visits, delivered in a health facility, and attended one postnatal visit. In the other 50% of census areas, pregnant women were not offered a conditional incentive.

Figure 2: Health care worker deployment (start month)

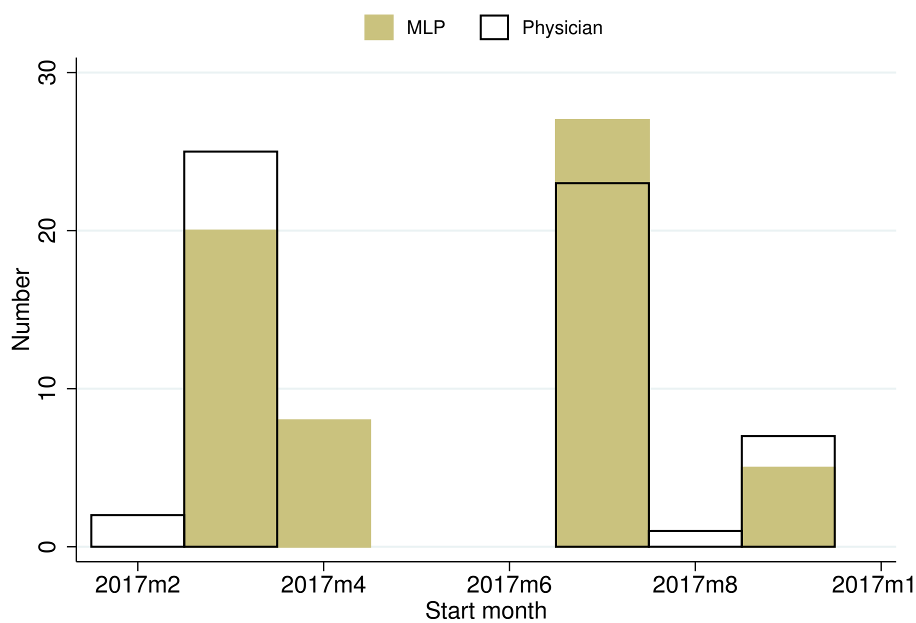


Figure 2 shows the distribution of health provider start months in clusters randomly assigned an additional health provider. 120 (out of 180) Primary Health Service Areas were randomly assigned an additional health provider: 60 were assigned a physician, and 60 were assigned a mid-level provider. 117 health providers were actually deployed: 57 physicians and 60 mid-level providers. Data is from administrative records.

Figure 3: Average number of health providers in the health center before/after provider deployment

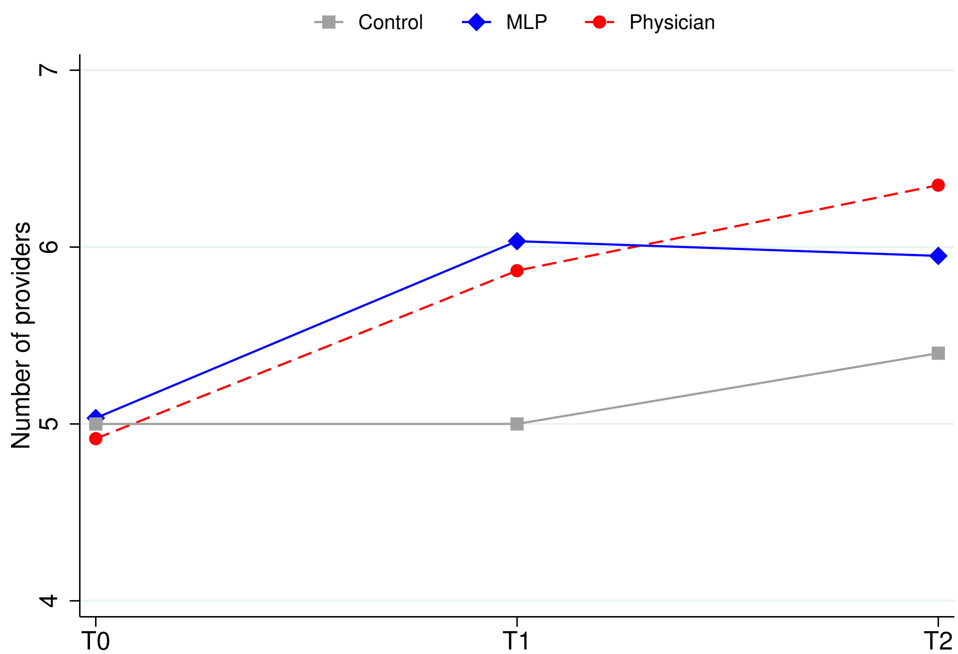


Figure 3 shows the mean number of health providers employed in the health center at each time point. Data is from health center staff registers. T0 is baseline, before health provider deployment; T1 is the beginning of the intervention period, just after the provider's expected start date; and T2 is the endline, just before the end of the provider's tenure. If the deployed providers resumed, the average number of health providers in the treated health center should increase by one. Control denotes health centers in clusters not assigned any additional health providers; MLP denotes health centers in clusters randomly assigned an additional mid-level health provider; Physician denotes health centers in clusters randomly assigned a physician provider.

Figure 4: Probability that there was a physician in the health center before/after deployment

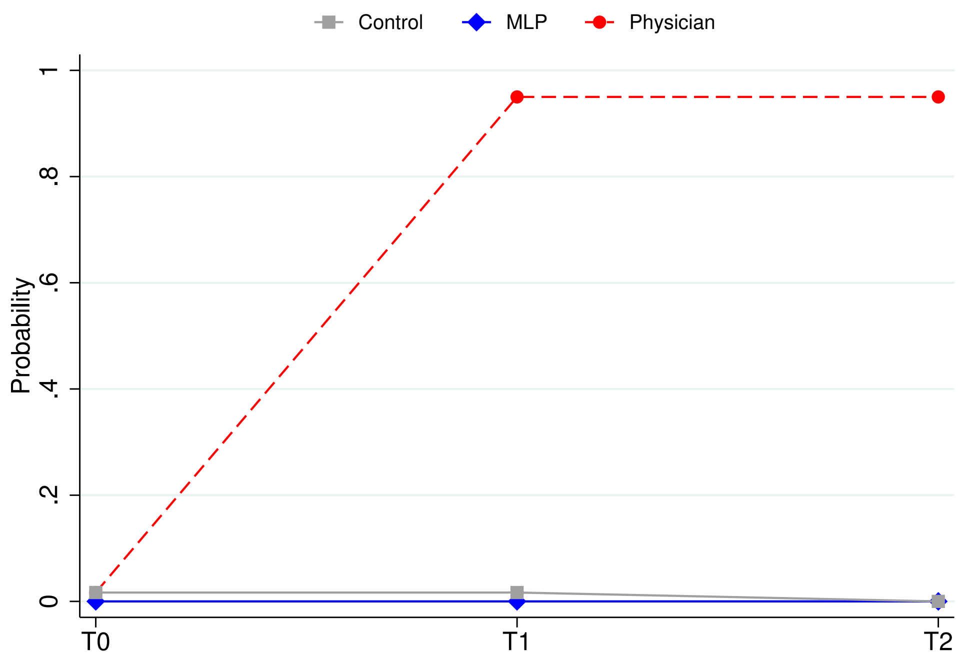


Figure 4 shows the probability that there was a physician present in the health center at each time point. Data is from health center staff registers. T0 is baseline, before health provider deployment; T1 is the beginning of the intervention period, just after the provider's expected start date; and T2 is the endline, just before the end of the provider's tenure. Control denotes health centers in clusters not assigned any additional health providers; MLP denotes health centers in clusters randomly assigned an additional mid-level health provider; Physician denotes health centers in clusters randomly assigned a physician provider.

Figure 5: Probability that the assigned health provider was present in the health center during their tenure

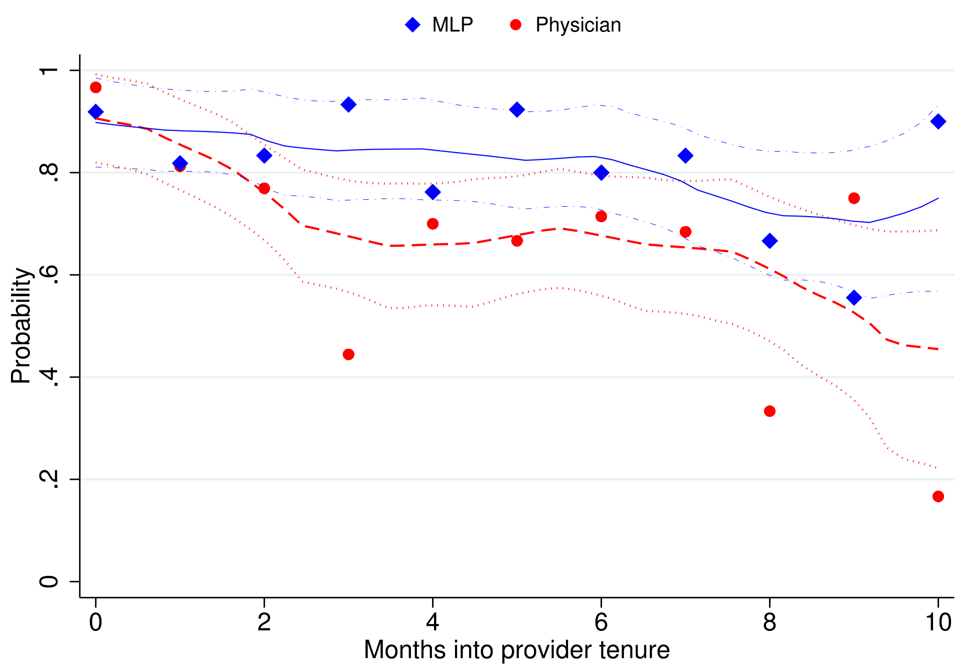
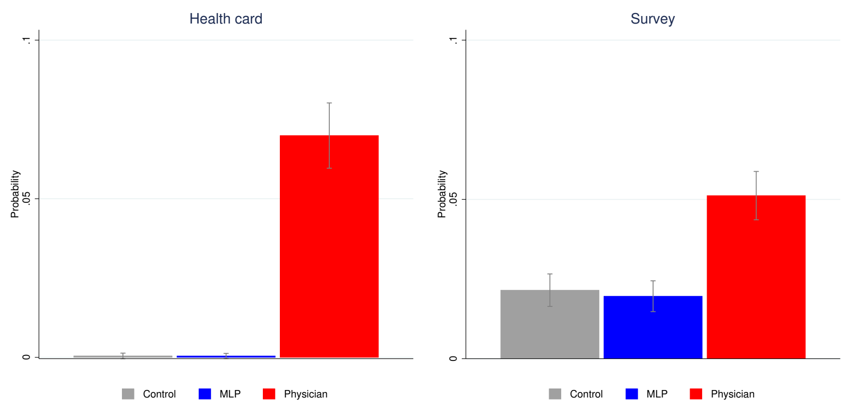
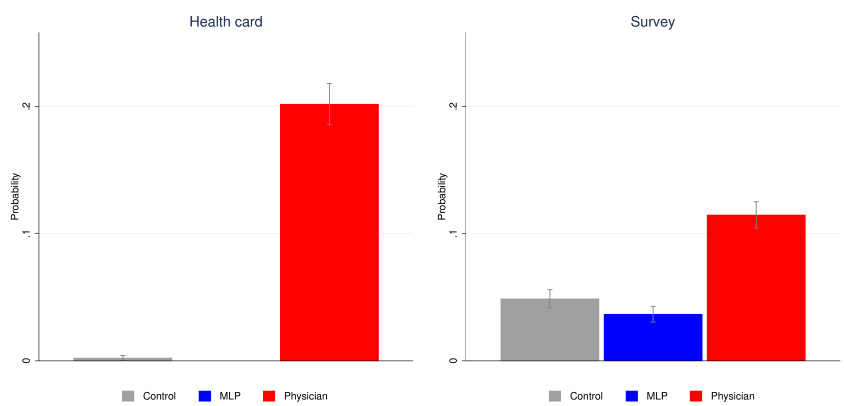


Figure 5 shows smoothed local polynomial regression lines and 95% confidence intervals. The dependent variable is the probability that the assigned health provider was physically present in the health center during surprise visits to the health center. Surprise visits were made to all health centers, including those not assigned an additional health provider. MLP denotes clusters randomly assigned an additional mid-level health provider; Physician denotes clusters randomly assigned a physician provider. Month 10 was the departure month for most providers.

Figure 6: Probability that medical care was provided by a physician by experimental arm



(a) Received prenatal care from a physician



(b) Received delivery care from a physician

Figure 6 shows the probability that medical care was provided by a physician during pregnancy (top figure), and delivery (bottom figure) by experimental arm. 95% confidence intervals shown. Data is from women's health cards (top and bottom left) or as reported in the endline women's survey (top and bottom right). Control denotes clusters not assigned any additional health providers; MLP denotes clusters randomly assigned an additional mid-level health provider; Physician denotes clusters randomly assigned a physician provider.

Figure 7: Means of 1-day and 7-day infant mortality by experimental arm

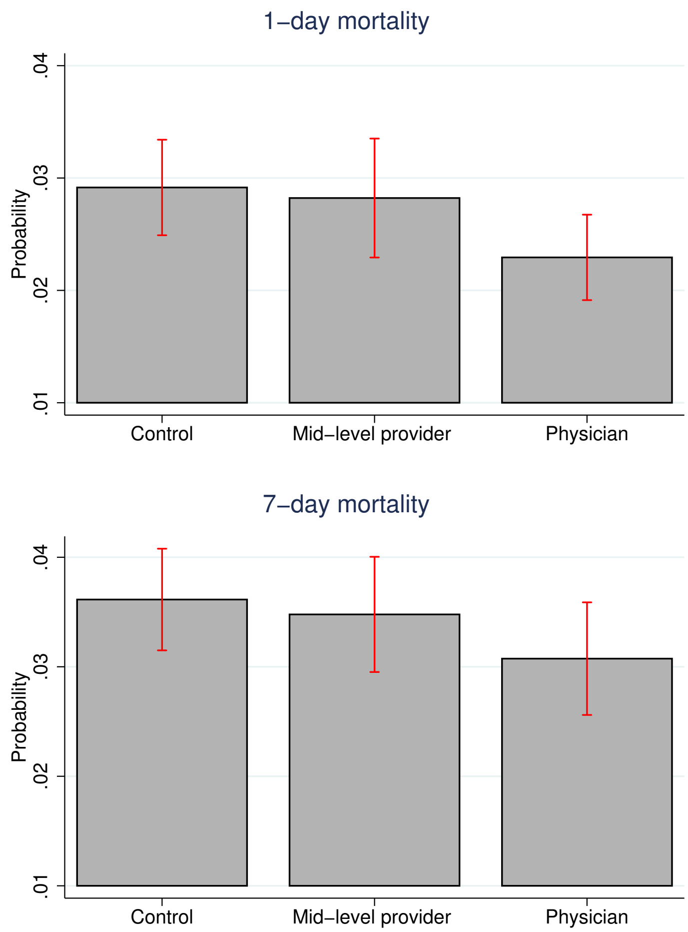


Figure 7 shows the probability that an infant died within 1 day of birth (top figure) or within 1 week of birth (bottom figure) by experimental arm. Data is from the endline women's survey. The results are from a regression of the outcome on the treatment assignment indicators and strata dummies. Control denotes clusters not assigned any additional health providers; MLP denotes clusters randomly assigned an additional mid-level health provider; Physician denotes clusters randomly assigned a physician provider.

Figure 8: Trends in 1-day and 7-day mortality by quarter of birth

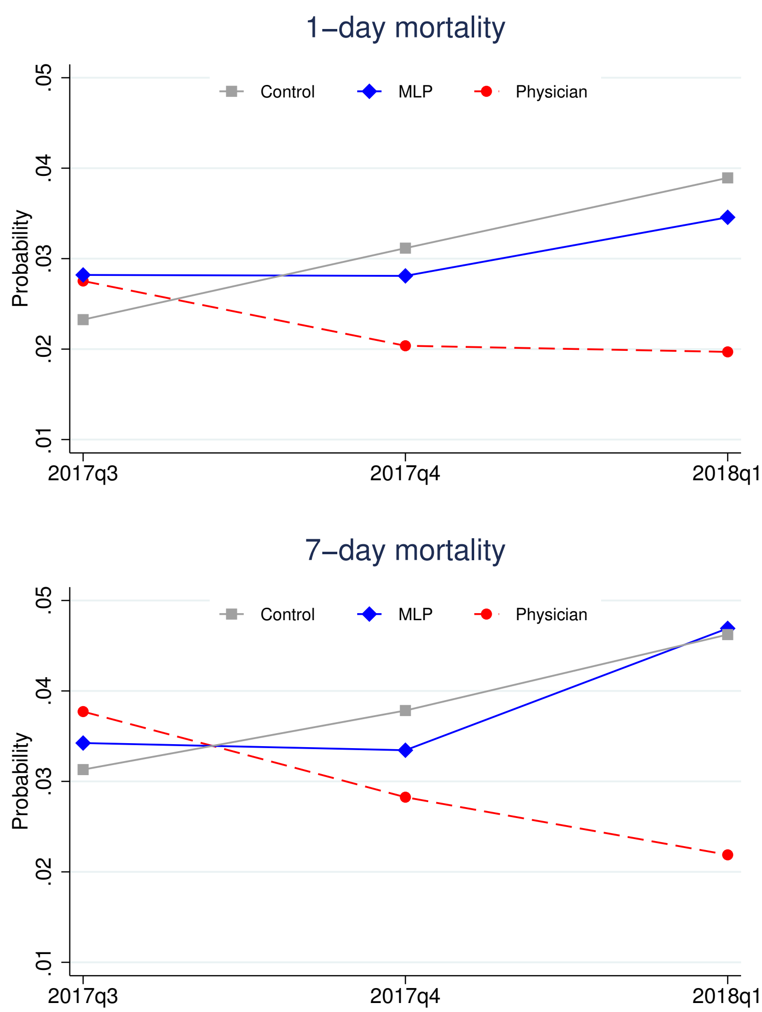
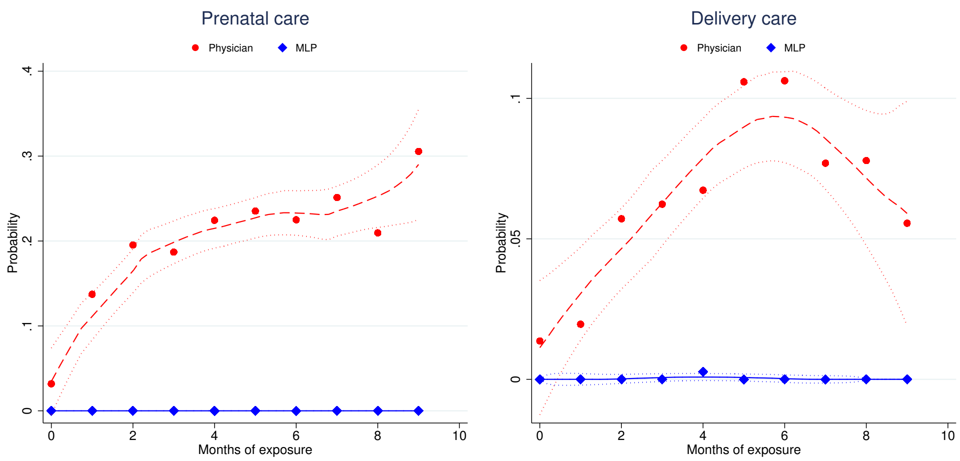
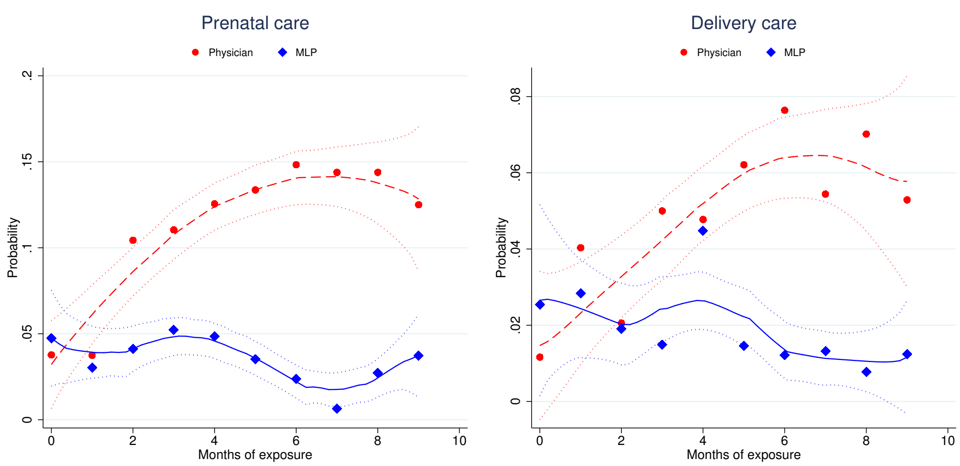


Figure 8 shows the probability that an infant died within 1 day of birth (top figure) or within 1 week of birth (bottom figure) by experimental arm and quarter of birth. Data is from the endline women's survey. 95% of births in the sample occurred within the period shown. Confidence intervals are omitted to reduce clutter. Control denotes clusters not assigned any additional health providers; MLP denotes clusters randomly assigned an additional mid-level health provider; Physician denotes clusters randomly assigned a physician provider.

Figure 9: Probability that medical care was provided by a physician by exposure duration



(a) Health card



(b) Survey

Figure 9 shows smoothed local polynomial regression lines and 95% confidence intervals. The dependent variable is the probability that medical care was provided by a physician during pregnancy (top and bottom left), and delivery (top and bottom right) by experimental arm and by exposure duration. Exposure is defined as the number of pregnancy months exposed to the intervention provider. Women who gave birth before the arrival of the new health care provider (or in the same month) are coded as having zero months of exposure. The maximum possible length of exposure is 10 months—the length of the provider's tenure. The probability that medical care was provided by a physician was calculated using information from women's health cards (Figure 9a) or as reported in the endline survey (Figure 9b). MLP denotes clusters randomly assigned an additional mid-level health provider; Physician denotes clusters randomly assigned a physician provider.

Figure 10: Means of 1-day and 7-day infant mortality by exposure duration

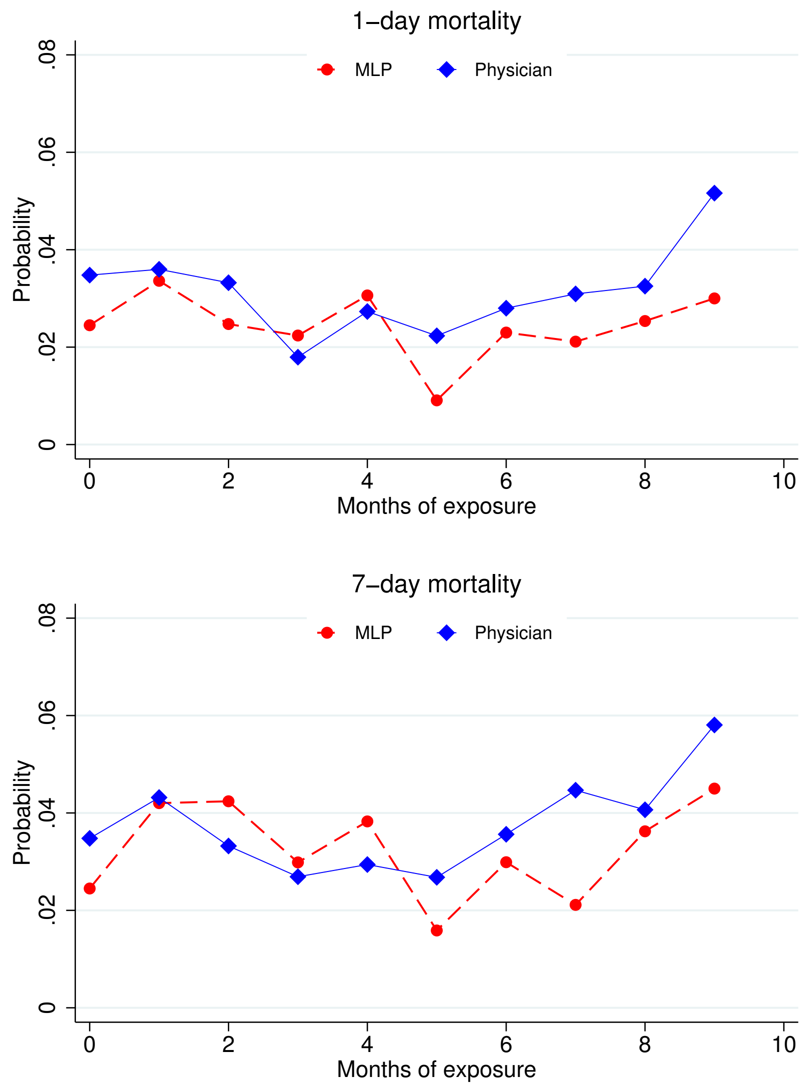


Figure 10 shows the probability that an infant died within 1 day of birth (top figure) or within 1 week of birth (bottom figure) by experimental arm and by exposure duration. Confidence intervals are omitted to reduce clutter. Exposure is defined as the number of pregnancy months exposed to the intervention provider. A woman who gave birth a month after the provider arrived is coded as having been exposed for 1 month. Women who gave birth before the arrival of the new health care provider, or in the same month, are coded as having zero months of exposure. The maximum possible length of exposure is 10 months—the length of the provider's tenure. MLP denotes clusters randomly assigned an additional mid-level health provider; Physician denotes clusters randomly assigned a physician provider.

Figure 11: Differences in clinical ability by provider type

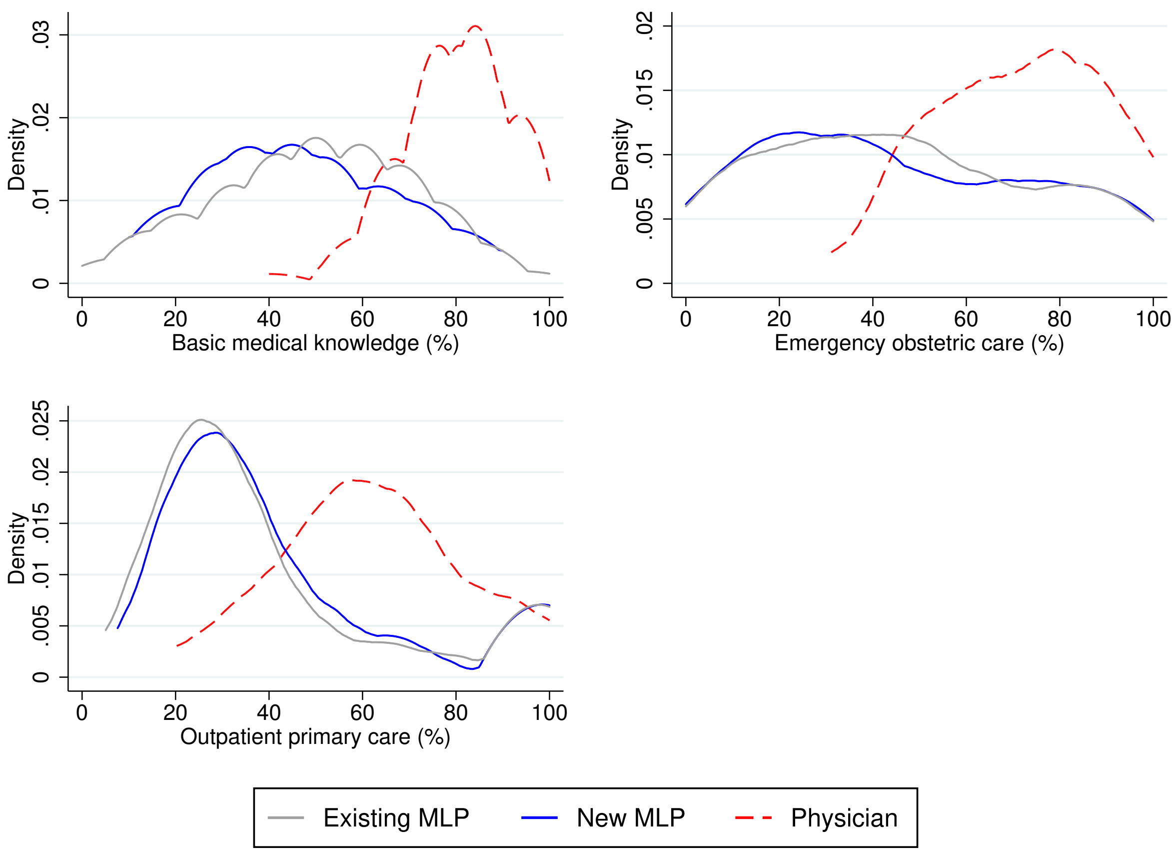
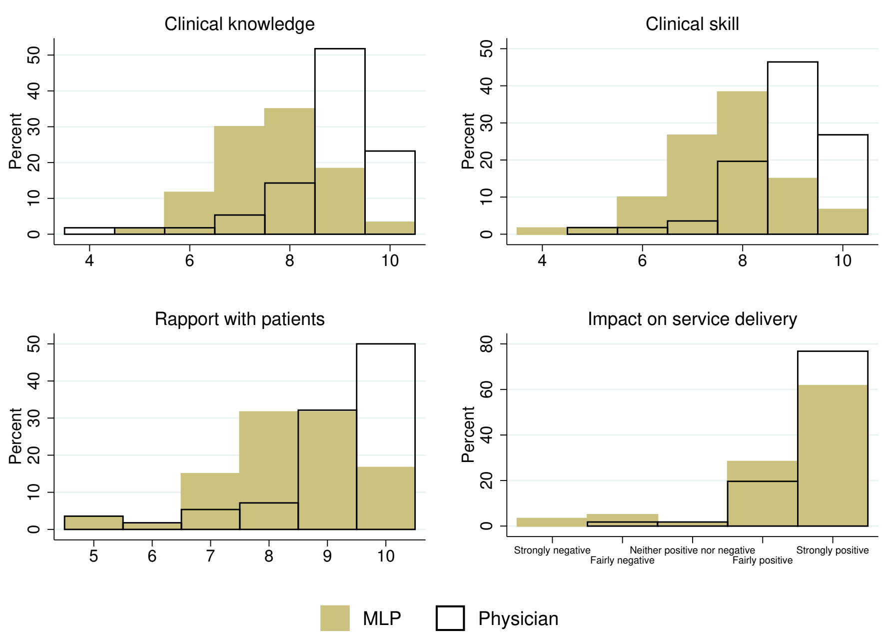


Figure 11 shows kernel density plots of health provider scores (out of 100) on clinical modules testing basic medical knowledge (top left), emergency obstetric case management (top right), and management of outpatient primary care conditions (bottom). The clinical modules were administered by medically trained professionals on the research team. MLP denotes a mid-level health provider. Figure 11 compares the new physicians to the new mid-level health providers, to existing mid-level providers in the health centers.

Figure 12: Assessments of assigned health care providers by health center managers



Confidential evaluations of each new health care provider were carried out by the clinic manager towards the end of the providers' tenure. Each clinic manager was asked to rate the performance of their assigned provider along various dimensions: their clinical knowledge, their clinical skill, and their rapport with patients. On each metric, the evaluator was asked to assign a score between 0 (the lowest) and 10 (the highest). Figure 12 shows the distribution of scores given by managers in physician-assigned health centers compared to managers in mid-level provider-assigned health centers. Managers were also asked to rate the new provider's impact on service delivery in the health center using a Likert scale. Results are shown bottom right.

Table 1: Baseline covariate balance

	Control	CHW	Physician	MLP=C	P=C	P=MLP	Joint
A: Health center variables							
Has running water	0.483	0.450	0.450	0.82	0.92	0.89	0.97
No electricity	0.267	0.283	0.267	0.71	0.99	0.67	0.90
Number of beds	16.500	14.233	15.117	0.14	0.49	0.34	0.32
Travel time to referral hospital (mins)	50.333	52.017	52.783	0.69	0.72	0.96	0.91
24-hour services	0.783	0.767	0.783	0.71	0.67	0.43	0.73
Inpatient care	0.810	0.702	0.764	0.15	0.53	0.45	0.36
Essential drugs (% in stock)	0.690	0.720	0.732	0.25	0.09	0.82	0.23
Essential equipment (% available)	0.541	0.550	0.515	0.71	0.58	0.30	0.58
Deliveries per month	24.894	24.861	23.264	0.98	0.59	0.68	0.85
Number of health care providers	5.000	5.017	4.917	0.75	0.68	0.92	0.91
Doctor	0.017	0.000	0.017	0.38	1.00	0.29	0.45
Nurse	0.450	0.333	0.367	0.34	0.43	0.86	0.60
Community Health Officer	0.383	0.317	0.317	0.50	0.59	0.87	0.79
Health Extension Worker	1.883	2.133	1.967	0.50	0.73	0.66	0.79
Junior Health Extension Worker	0.817	0.983	1.067	0.46	0.18	0.55	0.38
Other cadre	1.450	1.250	1.183	0.37	0.20	0.69	0.43
Sample size	60	60	60				
Omnibus test (p-value)							0.19
B: Mother variables							
Age	24.767	24.825	24.597	0.78	0.32	0.24	0.46
Hausa/Fulani ethnicity	0.737	0.700	0.769	0.13	0.39	0.01	0.02
Moslem	0.830	0.806	0.818	0.01	0.26	0.37	0.03
No formal schooling	0.710	0.700	0.697	0.31	0.50	0.81	0.59
Cannot read	0.731	0.769	0.756	0.16	0.31	0.86	0.36
Husband makes health-care decisions	0.656	0.644	0.678	0.83	0.09	0.09	0.17
Number of prior births	1.907	1.856	1.935	0.41	0.89	0.33	0.55
Prior stillbirth or newborn death	0.068	0.058	0.061	0.23	0.55	0.55	0.49
Months pregnant at enrollment	4.302	4.230	4.190	0.34	0.11	0.37	0.28
Offered conditional incentive	0.560	0.538	0.529	0.29	0.12	0.68	0.29
Household assets (out of 11)	2.063	2.045	1.978	0.95	0.49	0.47	0.72
Household size	5.847	5.819	5.485	0.52	0.10	0.23	0.22
Distance to health center (km)	6.417	6.109	5.492	0.96	0.88	0.93	0.99
Sample size	3467	3511	3608				
Omnibus test (p-value)							0.18
C: Child variables							
Male infant	0.522	0.532	0.532	0.28	0.30	0.89	0.47
Multiple birth	0.022	0.022	0.023	0.95	0.90	0.94	0.99
Caesarean delivery	0.007	0.008	0.004	0.53	0.28	0.07	0.19
Health card available	0.670	0.695	0.676	0.57	0.62	0.31	0.59
Sample size	3007	3025	3094				
Omnibus test (p-value)							0.88

Control (C) denotes clusters not assigned any additional health providers; MLP denotes clusters randomly assigned an additional mid-level provider; Physician (P) denotes clusters randomly assigned a physician. In Panel A each observation is a health center; N=180 (from a survey of health centers in participating clusters). In Panel B each observation is a woman with complete data (baseline and endline); N=10,856 (variables are from the baseline survey). In Panel C each observation is a liveborn child; N=9,126 (variables are from the endline survey). The figures in Columns 4-6 are p-values from a test of difference in group means. Column 7 is the p-value from a joint test of equality. P-values are adjusted for clustering.

Table 2: Did the intervention increase health provider supply?

	Number of health workers		Physician available	
	(1)	(2)	(3)	(4)
Mid-level provider	1.033**	0.917**	-0.017	-0.013
	(0.504)	(0.386)	(0.017)	(0.020)
Physician	0.867*	0.787**	0.933***	0.940***
	(0.473)	(0.377)	(0.033)	(0.031)
T2	0.400**	0.400**	-0.017	-0.017
	(0.174)	(0.185)	(0.017)	(0.018)
Mid-level provider x T2	-0.483*	-0.483*	0.017	0.017
	(0.251)	(0.268)	(0.017)	(0.018)
Physician x T2	0.083	0.083	0.017	0.017
	(0.278)	(0.296)	(0.029)	(0.031)
Observations	360	360	360	360
Control group mean	5.200	5.200	0.008	0.008
p-value (Physician = MLP)	0.742	0.734	0.000	0.000

The dependent variables are in the column headers. They are the number of health care providers working in the health center and the probability that the health center has a physician on staff. Each health center, in each of the 180 project clusters, was observed twice: at the beginning of the intervention (T1) and just before the end (T2). The arrival of the assigned health care worker marks the start of the intervention, and their departure marks its end. MLP denotes clusters randomly assigned an additional mid-level provider; Physician denotes clusters randomly assigned a physician provider. The omitted comparison group consists of clusters not assigned any additional health providers. Standard errors in parentheses are adjusted for clustering. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$..

Table 3: Intention-to-treat effect on probability that care was provided by a physician

(a) Health card

	Prenatal care		Delivery care	
	(1)	(2)	(3)	(4)
Mid-level provider	0.002	0.001	0.001	0.001
	(0.015)	(0.015)	(0.008)	(0.008)
Physician	0.205***	0.206***	0.071***	0.071***
	(0.027)	(0.027)	(0.013)	(0.013)
Controls	No	Yes	No	Yes
Observations	6891	6891	6891	6891
Control group mean	0.002	0.002	0.000	0.000
p-value (Physician = MLP)	0.000	0.000	0.000	0.000

(b) Survey data

	Prenatal care		Delivery care	
	(1)	(2)	(3)	(4)
Mid-level provider	-0.006	-0.005	0.002	0.003
	(0.012)	(0.012)	(0.006)	(0.006)
Physician	0.071***	0.072***	0.033***	0.033***
	(0.015)	(0.015)	(0.008)	(0.008)
Controls	No	Yes	No	Yes
Observations	10586	10586	9410	9410
Control group mean	0.048	0.048	0.020	0.020
p-value (Physician = MLP)	0.000	0.000	0.000	0.000

The dependent variables are in the table headers. Prenatal care denotes whether a woman was seen by a physician at least once during the prenatal period. Delivery care denotes whether a physician attended the delivery. This is based on information recorded on their health cards (Panel A) or as reported in the survey (Panel B). MLP denotes clusters randomly assigned an additional mid-level provider; Physician denotes clusters randomly assigned a physician provider. The omitted comparison group is clusters not assigned any additional health providers. In the covariate-adjusted models we control for the woman's age, ethnicity, religion, literacy level (based on whether she could read a simple sentence shown to her in English), pregnancy risk, whether she was offered a cash incentive, decision-making authority in the household (a dummy denoting whether the spouse is the sole decision-maker), and the number of assets owned by the household. Standard errors in parentheses are clustered at the level of the primary health service area. There are 180 clusters. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

Table 4: Intention-to-treat effect on infant mortality

(a) 1-day mortality

	Per 1000 live births			Per 1000 women
	(1)	(2)	(3)	(4)
Mid-level provider	-0.941	-1.261	-1.796	-1.611
	(3.438)	(3.425)	(3.427)	(2.994)
Physician	-6.224**	-6.866**	-8.168***	-6.448**
	(2.967)	(3.004)	(3.098)	(2.731)
Controls	No	No	Yes	Yes
Flexible time trends	No	Yes	Yes	No
Observations	9126	9125	9125	10586
Control group mean	28.932	28.942	28.942	25.094
p-value (Physician = MLP)	0.118	0.091	0.059	0.097

(b) 7-day mortality

	Per 1000 live births			Per 1000 women
	(1)	(2)	(3)	(4)
Mid-level provider	-1.359	-1.718	-2.580	-2.205
	(3.573)	(3.592)	(3.646)	(3.154)
Physician	-5.397	-6.096*	-7.759**	-6.050*
	(3.593)	(3.614)	(3.666)	(3.172)
Controls	No	No	Yes	Yes
Flexible time trends	No	Yes	Yes	No
Observations	9126	9125	9125	10586
Control group mean	36.249	36.261	36.261	31.439
p-value (Physician = MLP)	0.282	0.240	0.170	0.229

The dependent variables are indicators denoting an infant death within 1 day of birth (Table 4a), and within 1 week of birth (Table 4b). MLP denotes clusters randomly assigned an additional mid-level provider; Physician denotes clusters randomly assigned a physician provider. The omitted comparison group is clusters not assigned any additional health providers. In Column 2, we include state $\times$ birth year-month trends. In the covariate-adjusted models we control for the woman's age, ethnicity, religion, literacy level (based on whether she could read a simple sentence shown to her in English), pregnancy risk, whether she was offered a cash incentive, decision-making authority in the household (a dummy denoting whether the spouse is the sole decision-maker), and the number of assets owned by the household. We also control for birth characteristics including indicators for a first birth, male infant, and a multiple delivery. Standard errors in parentheses are clustered at the level of the primary health service area. There are 180 clusters. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 5: Intention-to-treat effect on infant mortality by treatment dosage

(a) 1-day mortality

	Low dose of treatment				High dose of treatment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Physician	0.436	0.174	-0.789	-1.136	-10.213**	-9.988**	-11.604**	-9.372**
	(5.410)	(5.467)	(5.857)	(4.404)	(4.206)	(4.183)	(4.441)	(4.007)
Controls	No	No	Yes	Yes	No	No	Yes	Yes
Flexible time trends	No	Yes	Yes	No	No	Yes	Yes	No
Observations	2916	2916	2916	3703	3201	3201	3201	3416

(b) 7-day mortality

	Low dose of treatment				High dose of treatment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Physician	4.535	4.258	3.650	1.968	-11.779**	-11.816**	-13.109**	-10.404**
	(6.183)	(6.185)	(6.653)	(4.979)	(5.202)	(5.204)	(5.581)	(5.074)
Controls	No	No	Yes	Yes	No	No	Yes	Yes
Flexible time trends	No	Yes	Yes	No	No	Yes	Yes	No
Observations	2916	2916	2916	3703	3201	3201	3201	3416

This table disaggregates effects by duration of exposure to the treatment (treatment dosage). Exposure is defined as the number of pregnancy months exposed to the intervention provider. Physician clusters are compared to mid-level provider clusters. Low dosage denotes women exposed for less than the median duration of 4 months. High dosage denotes exposure duration greater than the median. The dependent variables are indicators denoting an infant death within 1-day of birth (Table 5a), and within 1 week of birth (Table 5b). In Columns 1-3 and 5-7, coefficients are deaths per 1000 live births and in Columns 4 and 8, it is deaths per 1000 women. Physician denotes clusters randomly assigned a physician provider. In the covariate-adjusted models we control for the woman's age, ethnicity, religion, literacy level (based on whether she could read a simple sentence shown to her in English), pregnancy risk, whether she was offered a cash incentive, decision-making authority in the household (a dummy denoting whether the spouse is the sole decision-maker), and the number of assets owned by the household. We also control for birth characteristics including indicators for a first birth, male infant, and a multiple delivery. Flexible time trends denote state $\times$ birth year-month trends. Standard errors in parentheses are clustered at the level of the primary health service area. There are 180 clusters. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 6: Intention-to-treat effect on other mortality outcomes

	In utero child death		Perinatal deaths per 1000		Newborn deaths per 1000		Overall child survival	
	(1) Low dose	(2) High dose	(3) Low dose	(4) High dose	(5) Low dose	(6) High dose	(7) Low dose	(8) High dose
Physician	0.008	-0.008	11.286	-19.830**	5.665	-16.222**	-0.002	0.027***
	(0.007)	(0.006)	(9.991)	(8.040)	(7.240)	(6.253)	(0.010)	(0.009)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3046	3357	3046	3357	2916	3201	3046	3357
Control group mean	0.142	0.142	69.961	69.961	46.891	46.891	0.811	0.811

The dependent variables are in the table header. We examine the effects on the probability of an in utero child death (defined as a miscarriage or stillbirth), perinatal deaths (a stillbirth or an early newborn deaths), newborn deaths (death of a liveborn infant between day 0 and day 28), and the probability of overall child survival (the probability that a child who was in utero at enrollment was alive at the endline). We disaggregate the results by treatment dosage. Physician denotes clusters randomly assigned a physician provider. MLP denotes clusters randomly assigned an additional mid-level provider. In the covariate-adjusted models we control for the woman's age, ethnicity, religion, literacy level (based on whether she could read a simple sentence shown to her in English), pregnancy risk, whether she was offered a cash incentive, decision-making authority in the household (a dummy denoting whether the spouse is the sole decision-maker), and the number of assets owned by the household. Standard errors in parentheses are clustered at the level of the primary health service area. There are 180 clusters. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

Table 7: Sub-group analysis

(a) 1-day child mortality

	Child sex		Mother can read		Pregnancy risk factor		Offered incentive		Top quartile health center	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Girls	Boys	No	Yes	No	Yes	No	Yes	No	Yes
Mid-level provider	1.545	-6.642	-2.173	-4.669	-6.191	3.972	5.328	-7.582	-2.168	5.075
	(4.792)	(4.949)	(4.075)	(6.977)	(4.451)	(5.492)	(5.879)	(4.644)	(3.791)	(12.386)
Physician	-5.097	-12.337***	-7.877**	-9.610	-10.526***	-5.764	-7.685	-8.111*	-9.018**	-5.731
	(4.644)	(4.474)	(3.983)	(5.954)	(3.901)	(5.232)	(5.419)	(4.453)	(3.512)	(7.724)
Observations	4301	4824	6847	2278	5186	3939	3982	5143	7458	1667

(b) 7-day child mortality

	Child sex		Mother can read		Pregnancy risk factor		Offered incentive		Top quartile health center	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Girls	Boys	No	Yes	No	Yes	No	Yes	No	Yes
Mid-level provider	-4.234	-3.877	-2.962	-6.153	-6.645	2.276	4.822	-8.815*	-3.340	3.411
	(5.292)	(6.099)	(4.006)	(8.393)	(5.018)	(5.749)	(6.644)	(5.099)	(4.123)	(12.358)
Physician	-1.826	-14.359***	-5.550	-20.108***	-10.162**	-5.675	-4.485	-9.778**	-8.418**	-7.386
	(5.780)	(5.365)	(4.453)	(7.374)	(4.557)	(6.411)	(6.259)	(4.911)	(4.031)	(8.876)
Observations	4301	4824	6847	2278	5186	3939	3982	5143	7458	1667

We disaggregate mortality effects for various sub groups. The dependent variables are indicators denoting an infant death within 1 day of birth (Table 7a), and within 1 week of birth (Table 7b). MLP denotes clusters randomly assigned an additional mid-level provider; Physician denotes clusters randomly assigned a physician provider. The omitted comparison group is clusters not assigned any additional health providers. All models include controls for the woman's age, ethnicity, religion, literacy level (based on whether she could read a simple sentence shown to her in English), pregnancy risk, whether she was offered a cash incentive, decision-making authority in the household (a dummy denoting whether the spouse is the sole decision-maker), and the number of assets owned by the household. We also control for birth characteristics including indicators for a first birth, male infant, a multiple delivery, and state × birth year-month trends. Standard errors in parentheses are clustered at the level of the primary health service area. There are 180 clusters. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Table 8: Did substituting a physician for a mid-level provider lead to improvements in quality?

	Good clinical practice					Bad clinical practice	
	(1) Adherence to fever protocol	(2) Carried out physical exam	(3) Made a diagnosis	(4) Patient communication	(5) Log of consultation time	(6) Prescribed injection	(7) Prescribed antibiotic
New MLP	0.003	0.020	-0.059	-0.014	0.082*	0.003	0.001
	(0.023)	(0.034)	(0.041)	(0.018)	(0.047)	(0.031)	(0.034)
Physician	0.148***	0.079**	0.248***	0.052***	0.281***	-0.084***	-0.107***
	(0.025)	(0.033)	(0.033)	(0.019)	(0.052)	(0.028)	(0.036)
Observations	1169	2390	2390	2390	2383	2392	2392
Dep. variable mean	0.240	0.816	0.669	0.488	2.096	0.225	0.457

This table examines various metrics of outpatient care quality as measured by direct observation. The column headers indicate the various quality metrics. The model compares physicians to the new mid-level health providers and existing mid-level providers (the omitted comparison group). Each regression controls for the following patient characteristics: age, sex, illness severity, self-rated health, fever presentation, a pregnancy-related visit, the consultation order, whether the consultation was interrupted, and mode of transportation to the health center. Standard errors in parentheses are clustered at the level of the primary health service area. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

Table 9: Intention-to-treat effect on quality of obstetric care

	Received a uterotonic		Received cord traction	
	(1)	(2)	(3)	(4)
Mid-level provider	0.019	0.022	0.035	0.037
	(0.017)	(0.016)	(0.023)	(0.023)
Physician	0.037**	0.041***	0.046**	0.050**
	(0.016)	(0.015)	(0.021)	(0.021)
Controls	No	Yes	No	Yes
Observations	9521	9521	9521	9521
Control group mean	0.325	0.325	0.362	0.362
p-value (Physician = MLP)	0.298	0.231	0.644	0.560

We examine two indicators of high quality obstetric care. Uterotonics are drugs that cause the uterus to contract. In Columns 3-4, the dependent variable is an indicator denoting whether a woman received cord traction to deliver the placenta. MLP denotes clusters randomly assigned an additional mid-level provider; Physician denotes clusters randomly assigned a physician provider. The omitted comparison group consists of clusters assigned an additional mid-level provider. In the covariate-adjusted models we control for the woman's age, ethnicity, religion, literacy level (based on whether she could read a simple sentence shown to her in English), pregnancy risk, whether she was offered a cash incentive, decision-making authority in the household (a dummy denoting whether the spouse is the sole decision-maker), and the number of assets owned by the household. We also control for birth characteristics including indicators for a first birth, male infant, and a multiple delivery. Standard errors in parentheses are clustered at the level of the primary health service area. There are 180 clusters. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 10: Behavioral responses to the supply shock

(a) Prenatal care

	(1) Any prenatal	(2) Public hospital	(3) Health center	(4) Other public	(5) Private facility	(6) Other
Mid-level provider	0.024*	0.012	-0.013	0.013	0.001	0.011*
	(0.013)	(0.013)	(0.024)	(0.014)	(0.002)	(0.006)
Physician	-0.002	-0.014	0.023	-0.008	-0.002	0.004
	(0.014)	(0.011)	(0.024)	(0.013)	(0.002)	(0.007)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10586	10586	10586	10586	10586	10586
Control group mean	0.777	0.068	0.626	0.068	0.007	0.028
p-value (Physician = MLP)	0.068	0.046	0.143	0.112	0.116	0.319

(b) Place of delivery

	(1) At home	(2) Public hospital	(3) Health center	(4) Other public	(5) Private facility	(6) Other
Mid-level provider	-0.047***	0.003	0.030	0.009	0.001	0.004
	(0.018)	(0.006)	(0.019)	(0.006)	(0.003)	(0.006)
Physician	-0.047**	-0.007	0.048**	0.003	-0.003	0.005
	(0.018)	(0.006)	(0.019)	(0.005)	(0.002)	(0.006)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9410	9410	9410	9410	9410	9410
Control group mean	0.561	0.033	0.325	0.017	0.007	0.057
p-value (Physician = MLP)	0.990	0.064	0.373	0.271	0.153	0.789

In Panel A, we examine whether the intervention affected a woman's probability of using prenatal care (Column 1), or the location where she sought prenatal care (Columns 2-6). In Panel B, we examine whether the intervention affected a woman's probability of a home birth (Column 1), or the location where she sought delivery care (Columns 2-6). MLP denotes clusters randomly assigned an additional mid-level provider; Physician denotes clusters randomly assigned a physician provider. The omitted comparison group is clusters not assigned any additional health providers. All models include controls for the woman's age, ethnicity, religion, literacy level (based on whether she could read a simple sentence shown to her in English), pregnancy risk, whether she was offered a cash incentive, decision-making authority in the household (a dummy denoting whether the spouse is the sole decision-maker), and the number of assets owned by the household. Standard errors in parentheses are clustered at the level of the primary health service area. There are 180 clusters. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

Table 11: What are the returns to physician human capital?

(a) OLS Estimates

	1-day mortality		7-day mortality	
	(1)	(2)	(3)	(4)
Medical care	-0.003	-0.004	-0.003	-0.004
	(0.005)	(0.005)	(0.006)	(0.006)
Medical care × physician	-0.012***	-0.014***	-0.010*	-0.012**
	(0.004)	(0.004)	(0.005)	(0.006)
Controls	No	Yes	No	Yes
Observations	6209	6208	6209	6208
Full Physician effect	-0.016***	-0.018***	-0.013*	-0.016**

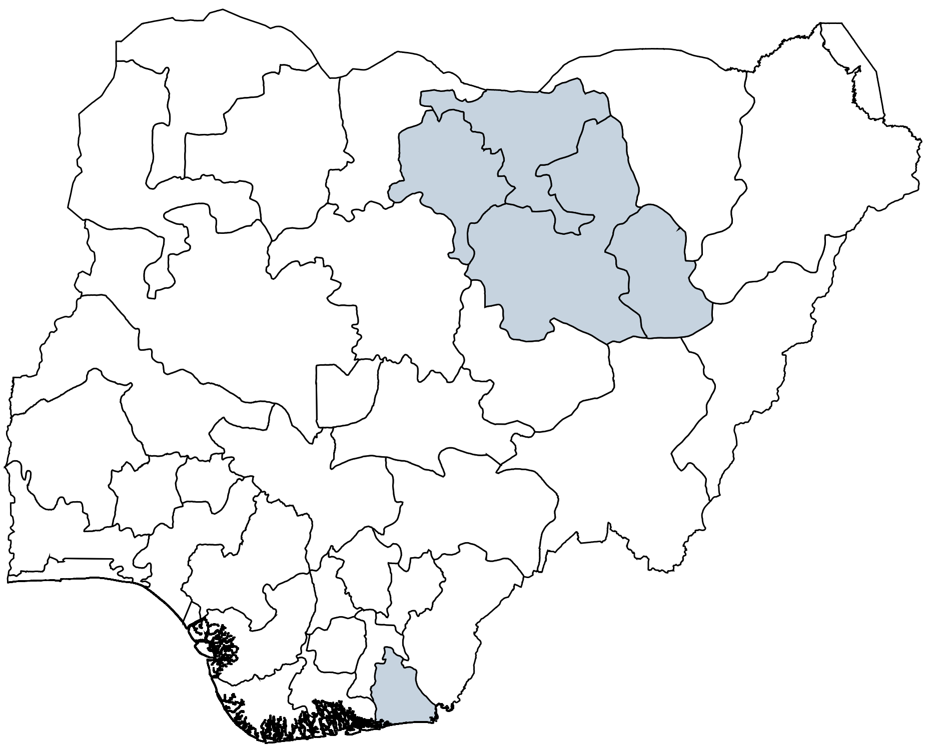
(b) IV Estimates

	(1)	(2)	(3)	(4)
Medical care	-0.028	-0.033	-0.045	-0.051
	(0.038)	(0.039)	(0.044)	(0.045)
Medical care × physician	-0.021	-0.028*	-0.029	-0.036*
	(0.015)	(0.016)	(0.019)	(0.020)
Controls	No	Yes	No	Yes
Observations	6209	6208	6209	6208
First-stage F-statistic	16.079	14.932	16.079	14.932
Full Physician effect	-0.049	-0.060	-0.073	-0.087*

The dependent variables are shown in the table header. 1-day (7-day) mortality denotes an infant death within 1 day (1 week) of birth. Medical care denotes women who attended at least three prenatal visits or gave birth in a health care facility. Physician-provided care is constructed using information from women's health cards. Coefficients are probabilities. Covariate-adjusted models include the same set of controls in Table 4 Column 3. Standard errors in parentheses are clustered at the level of the primary health service area. There are 180 clusters. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

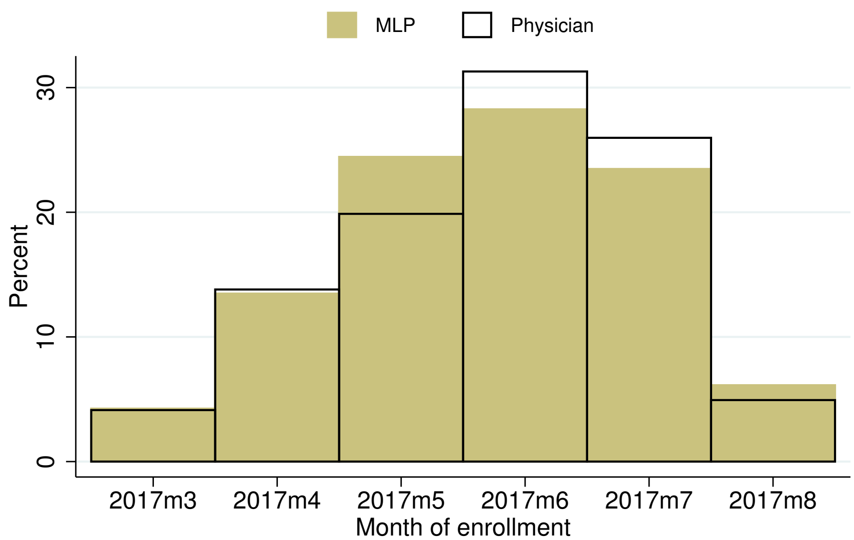
Appendix

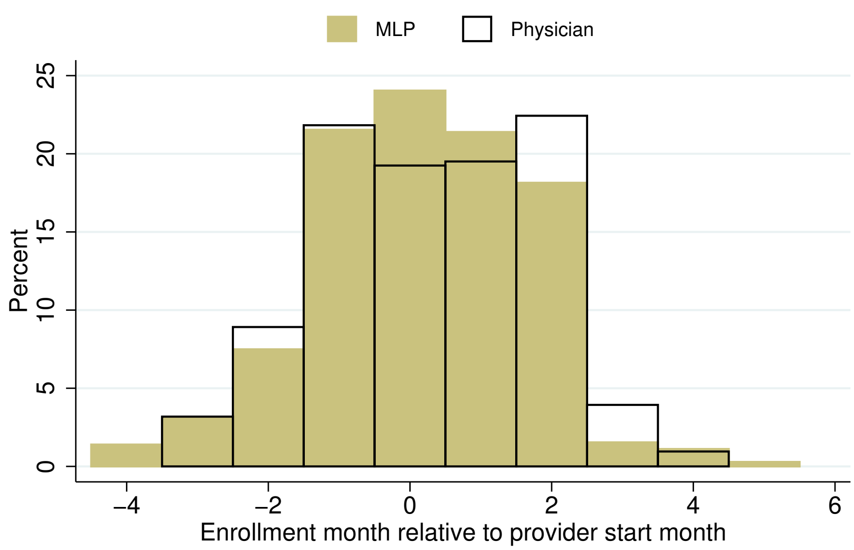
Figure A.1: Map of Nigeria showing Project States



The 180 project clusters were drawn from five states (shaded areas) representing three of Nigeria's six geopolitical regions. Up top from left to right: Kano, Jigawa, Bauchi, and Gombe. At the bottom is Akwa Ibom.

Figure A.2: Baseline enrollment





Top figure shows the period over which enrollment of households/women took place. MLP denotes clusters randomly assigned an additional mid-level provider; Physician denotes clusters randomly assigned a physician provider. Bottom figure shows enrollment relative to the arrival month of the new health provider. The month of arrival is month 0. $X < 0$ denotes enrollment X months prior to the arrival of the new provider. $X > 0$ denotes enrollment after the arrival of the new provider.

Figure A.3: Probability that delivery was recorded on woman's health card by delivery location and experimental arm

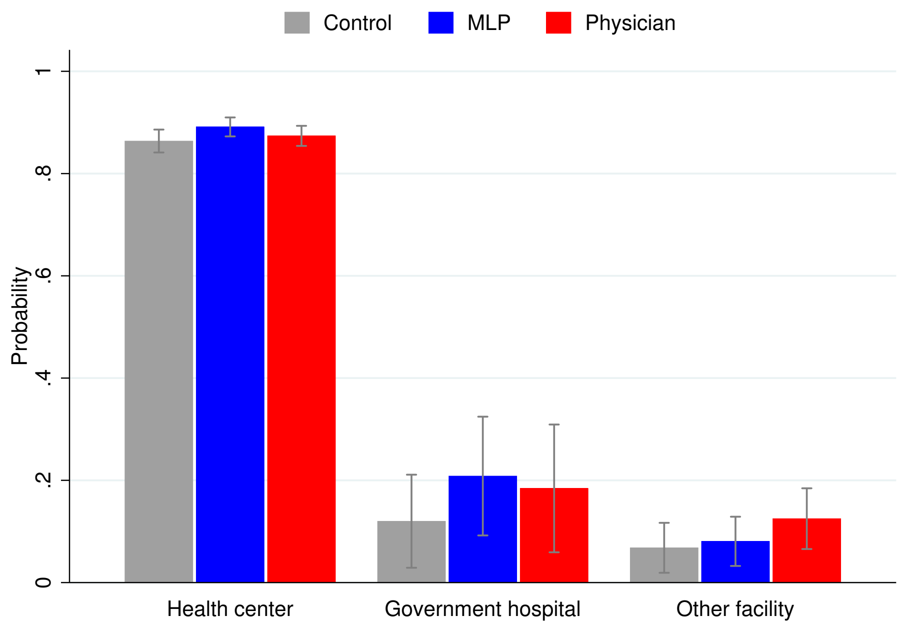


Figure A.3 shows the proportion of women whose delivery was recorded on their card by delivery location outside the home, and compares this across experimental arms. MLP denotes clusters randomly assigned an additional mid-level provider; Physician denotes clusters randomly assigned a physician provider. Other facility denotes births in any other location outside of the home. For reference, 77% of non-home births took place in the health center serving the cluster, 6% took place in a government hospital, and 17% took place in some other location (3.6% in another public facility, 1.2% in a private hospital or clinic, 6.4% in a maternity home, 4.5% in a church, and 1.2% elsewhere).

Figure A.4: Physician reported present at birth by delivery location

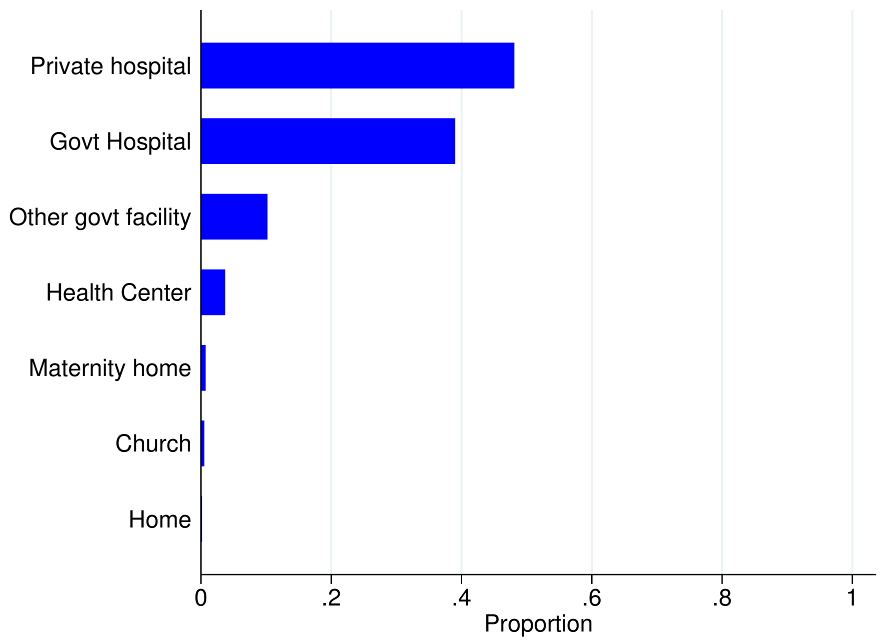


Figure A.4 shows the proportion of births attended by a physician by delivery location, as reported in the endline survey.

Figure A.5: Probability of a facility delivery by number of prenatal visits

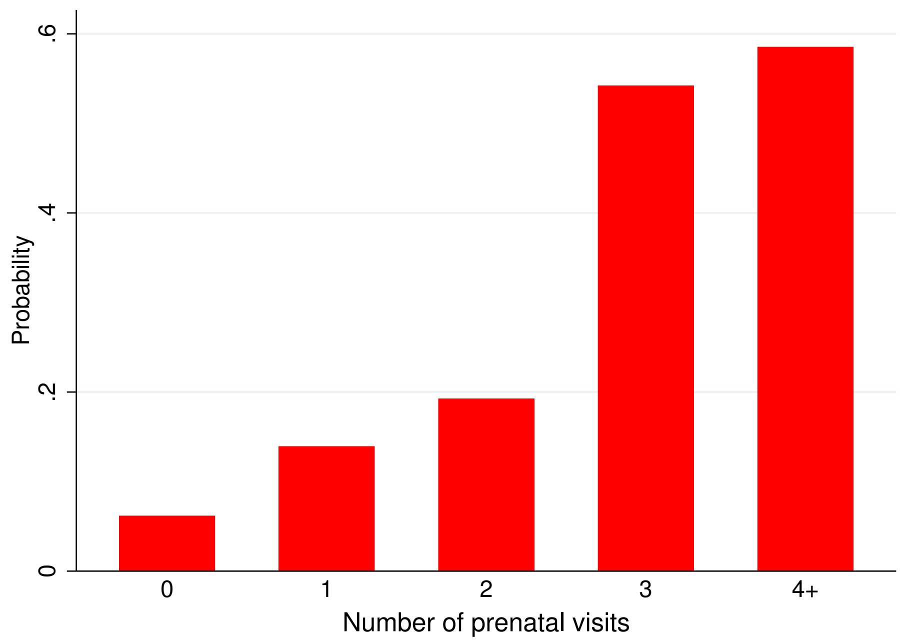


Figure A.5 shows the proportion of facility births by the number of prenatal care visits attended.

Figure A.6: Dosage of treatment: Number of pregnancy months of exposure

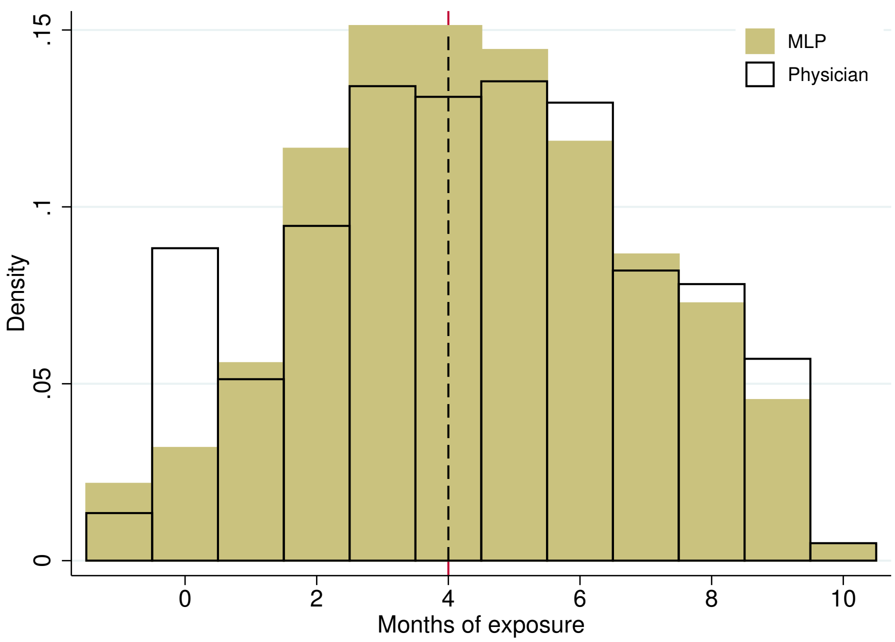


Figure A.6 shows the distribution of exposure duration in months. Exposure is defined as the number of pregnancy months exposed, based on the month when the pregnancy ended relative to the month of arrival of the health provider. A woman who gave birth a month after the provider arrived was exposed for 1 month. Physician denotes clusters randomly assigned a physician provider. MLP denotes clusters randomly assigned a mid-level provider provider. The peak at zero is because in clusters randomly assigned a physician, where one was not deployed, the number of exposure months is zero. The dotted line represents the median.

Figure A.7: Probability that medical care was provided by a physician by state

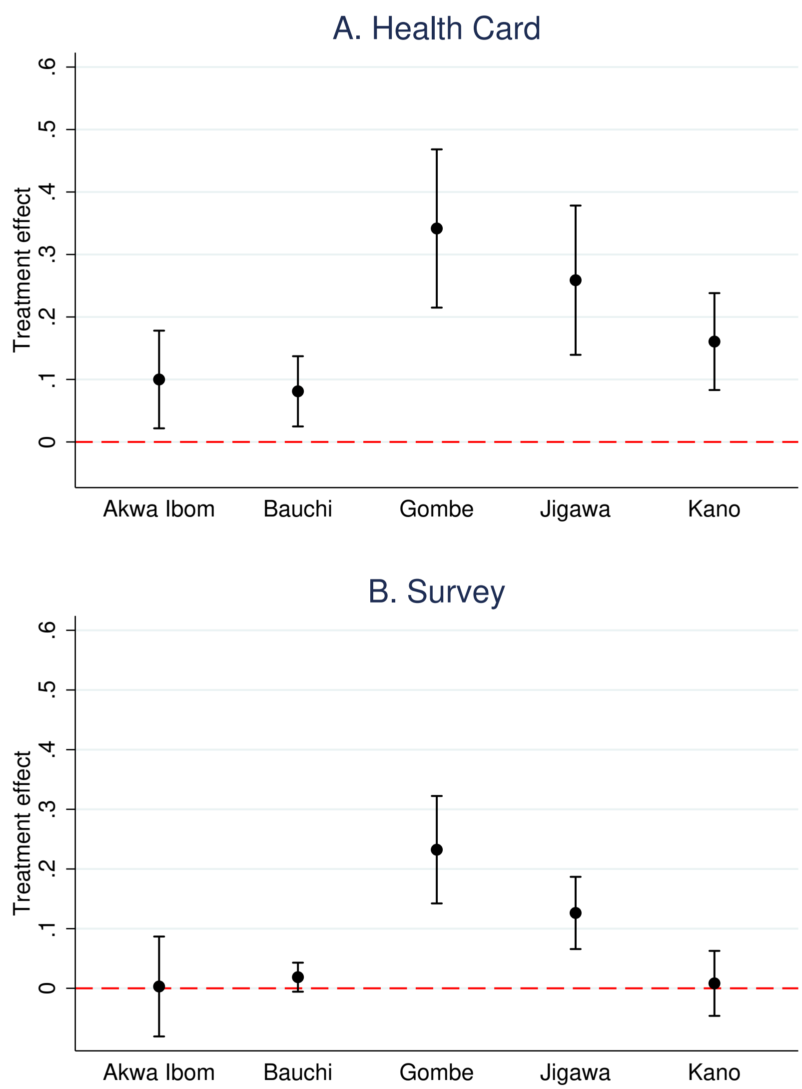


Figure A.7 shows the effect of being assigned a physician on the probability that a woman received care from a physician during pregnancy or delivery. We plot the point estimates and 95% confidence intervals from separate regressions for each state. In Figure A.7a, the dependent variable is constructed using data from women's health cards and, in Figure A.7b, using the survey data.

Figure A.8: Differences in performance by health provider qualifications

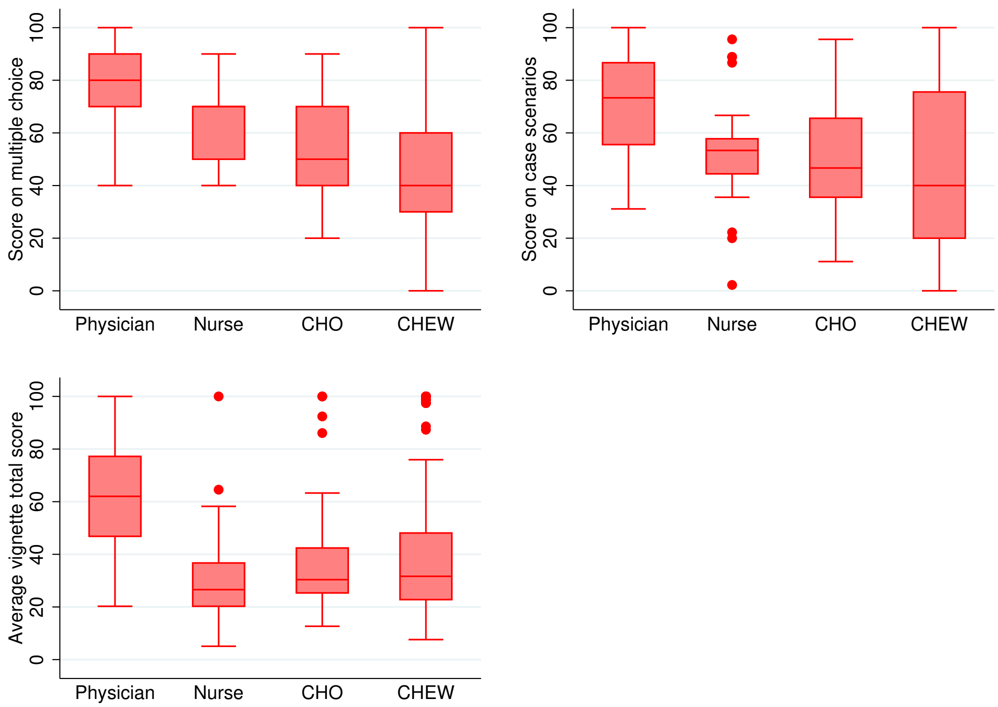


Figure 11 shows box plots of health provider scores (out of 100) on clinical modules testing basic medical knowledge (top left), emergency obstetric case management (top right), and management of outpatient primary care conditions (bottom). The clinical modules were administered by medically trained professionals on the research team. Figure 11 compares health care providers based on their medical qualifications. CHO denotes Community Health Officers, and CHEW denotes Community Health Extension Workers.

Figure A.9: Rates of correct diagnosis and treatment by provider type

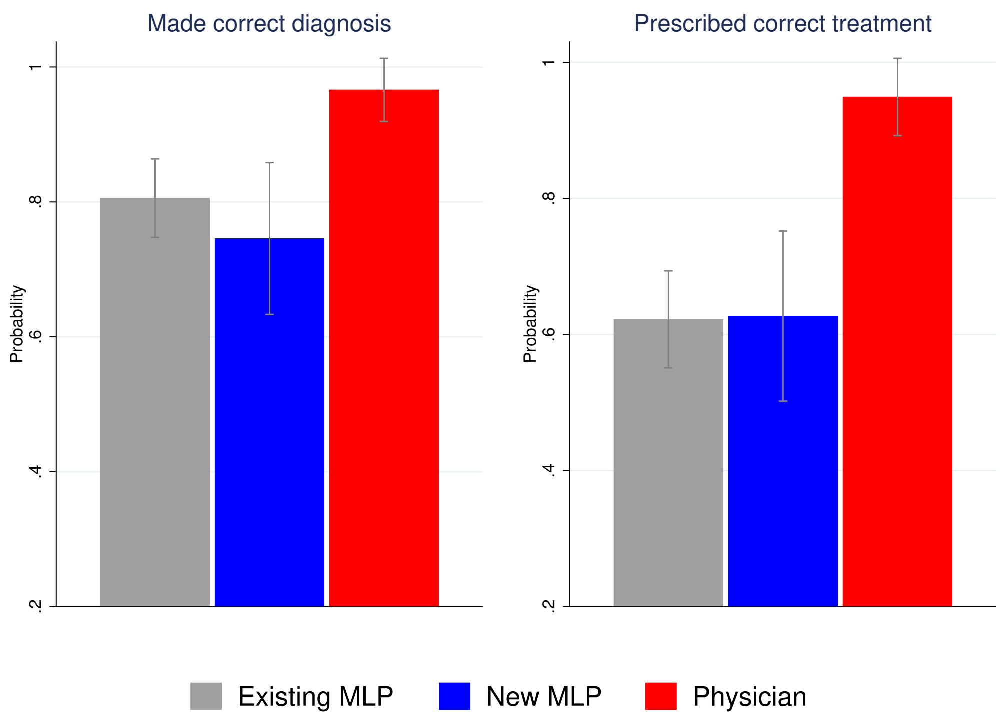


Figure A.9 examines rates of correct diagnosis and treatment for a case of tuberculosis presented using a patient vignette. We compare the new physicians to the new mid-level health providers, to existing mid-level providers in the health centers.

Figure A.10: Was there differential monitoring by experimental arm?

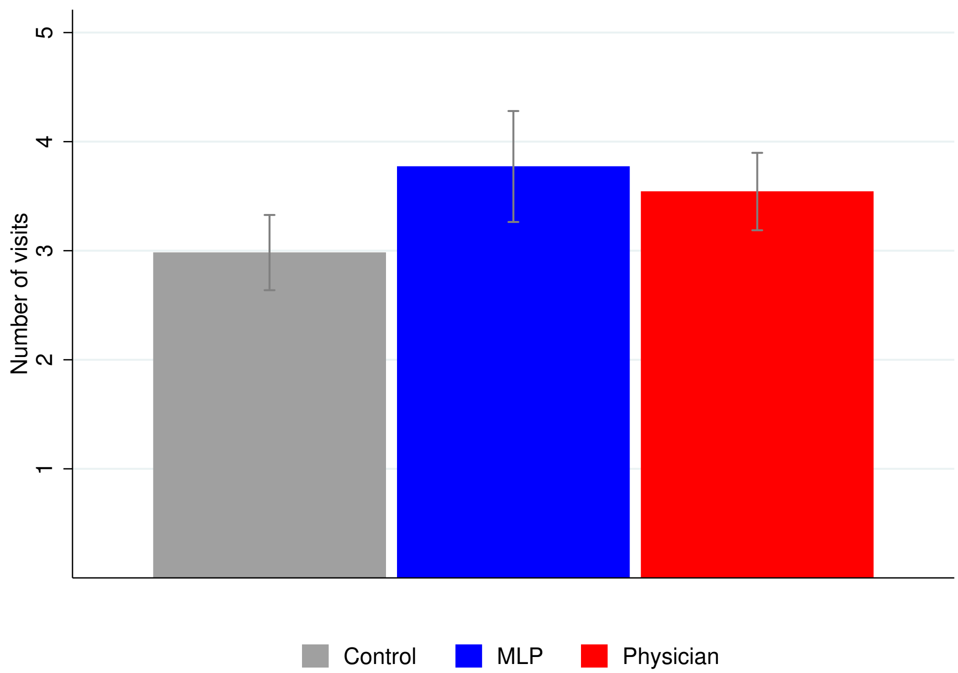
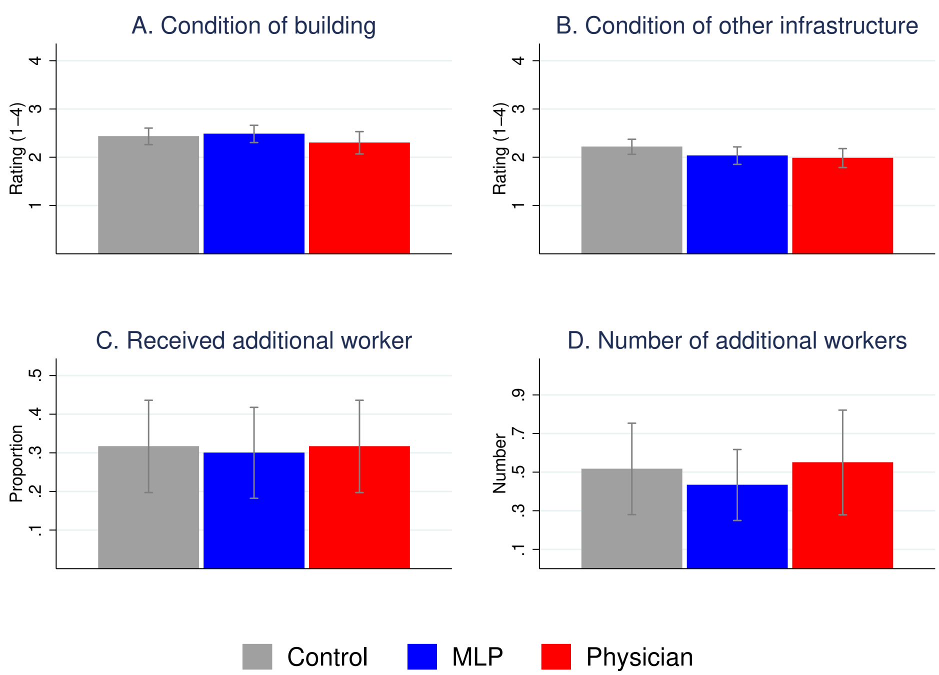


Figure A.10 shows the mean number of surprise visits by project staff to participating health centers in each arm of the experiment. MLP denotes clusters randomly assigned an additional mid-level provider; Physician denotes clusters randomly assigned a physician provider.

Figure A.11: Was there differential provision of human or capital resources to health centers?



During the endline visit to participating health centers, project staff observed and separately rated the condition of the building, and other health center infrastructure such as tables, chairs, beds, and screens. Infrastructure upgrades or additional capital expenditure would potentially show up here. Condition was rated on a four-point scale from one (poor) to four (excellent). Means and 95% confidence intervals by arm are shown in Figure A.11a and Figure A.11b. In Figure A.11c, the dependent variable is the probability that the health center received any additional workers between baseline and endline (excluding the deployed provider). In Figure A.11d, we plot the mean number of new workers by experimental arm. MLP denotes clusters randomly assigned an additional mid-level provider; Physician denotes clusters randomly assigned a physician provider.

Table A.1: Were attriters different from non-attriters?

	Non-attriters	Attriters	p-value
Mother variables			
Age	24.728	25.091	0.98
Hausa/Fulani ethnicity	0.736	0.485	0.30
Moslem	0.818	0.530	0.85
No formal schooling	0.702	0.492	0.03
Cannot read	0.752	0.632	0.01
Husband makes health-care decisions	0.660	0.575	0.89
Number of prior births	1.900	2.233	0.00
Prior stillbirth or newborn death	0.062	0.053	0.43
Months pregnant at enrollment	4.240	4.194	0.07
Offered conditional incentive	0.542	0.429	0.00
Household assets (out of 11)	2.028	2.504	0.27
Household size	5.714	4.996	0.71
Distance to health center (km)	6.000	4.550	0.67
Sample size	10586	266	
Omnibus test (p-value)			0.00

We compare the baseline characteristics of women who dropped out between baseline and endline (attriters) to women who did not (non-attriters). We cannot compare child characteristics because these variables are only in the endline survey. p-values are from a test of difference in group means and are adjusted for clustering.

Table A.2: Was there differential attrition?

	Control	MLP	Physician	MLP=C	P=C	P=MLP	Joint
Mother variables							
Age	25.110	24.849	25.259	0.78	0.84	0.97	0.95
Hausa/Fulani ethnicity	0.438	0.535	0.500	0.64	0.79	0.30	0.57
Moslem	0.486	0.563	0.556	0.53	0.71	0.91	0.82
No formal schooling	0.429	0.563	0.511	0.20	0.75	0.40	0.43
Cannot read	0.543	0.690	0.689	0.24	0.59	0.76	0.49
Husband makes health-care decisions	0.648	0.563	0.500	0.37	0.14	0.50	0.30
Number of prior births	2.171	1.887	2.578	0.11	0.53	0.05	0.10
Prior stillbirth or newborn death	0.057	0.028	0.067	0.47	0.76	0.41	0.58
Months pregnant at enrollment	4.371	4.242	3.954	0.82	0.55	0.37	0.67
Offered conditional incentive	0.467	0.394	0.411	0.55	0.99	0.62	0.81
Household assets (out of 11)	2.771	2.239	2.400	0.32	0.97	0.28	0.47
Household size	4.943	4.915	5.122	0.67	0.60	0.44	0.74
Distance to health center (km)	5.781	5.285	2.533	0.97	0.80	0.86	0.88
Sample size	105	71	90				
Omnibus test (p-value)							0.99

We compare the baseline characteristics of attriters by experimental arm. Control (C) denotes clusters not assigned any additional health providers; MLP denotes clusters randomly assigned an additional mid-level provider; Physician (P) denotes clusters randomly assigned a physician. The figures in Columns 4-6 are p-values from a test of difference in group means. Column 7 is the p-value from a joint test of equality. p-values are adjusted for clustering.

Table A.3: Robustness: Intention-to-treat effect on 1-day and 7-day infant mortality by state

	Gombe and Jigawa		Other three states	
	(1) 1-day	(2) 7-day	(3) 1-day	(4) 7-day
Mid-level provider	-2.402	-3.671	-0.324	-0.844
	(5.486)	(5.659)	(4.587)	(4.866)
Physician	-10.173**	-9.766**	-6.336	-5.675
	(4.110)	(4.302)	(4.594)	(5.647)
Observations	4117	4117	5008	5008
Control group mean	22.070	28.919	34.279	41.962
p-value (Physician = MLP)	0.112	0.260	0.218	0.368

The dependent variables are indicators denoting an infant death within 1 day, and 7 days of birth. We present results for the two states—Gombe and Jigawa—that saw the largest increase in physician-provided medical care vs. the three other states that saw much smaller increases (see Figure A.7). Coefficients are deaths per 1000 live births. MLP denotes clusters randomly assigned an additional mid-level provider; Physician denotes clusters randomly assigned a physician provider. The omitted comparison group is clusters not assigned any additional health providers. All models include the same set of controls in Table 4 Column 3. Standard errors in parentheses are clustered at the level of the primary health service area. There are 180 clusters. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

Table A.4: Intention-to-treat effect on birthweight

	Birthweight (kg)		Birthweight <2.5kg	
	(1)	(2)	(3)	(4)
Mid-level provider	-0.013	-0.008	-0.009	-0.013
	(0.032)	(0.033)	(0.014)	(0.013)
Physician	0.020	0.019	-0.021*	-0.022*
	(0.032)	(0.030)	(0.011)	(0.012)
Controls	No	Yes	No	Yes
Observations	2073	2073	2073	2073
Control group mean	3.135	3.135	0.078	0.078
p-value (Physician = MLP)	0.359	0.454	0.345	0.455

The dependent variables are shown in the table header. Birthweight data are only available for a subset of infants. MLP denotes clusters randomly assigned an additional mid-level provider; Physician denotes clusters randomly assigned a physician provider. In the covariate-adjusted models we control for the woman's age, ethnicity, religion, literacy level (based on whether she could read a simple sentence shown to her in English), pregnancy risk, whether she was offered a cash incentive, decision-making authority in the household (a dummy denoting whether the spouse is the sole decision-maker), and the number of assets owned by the household. We also control for birth characteristics including indicators for a first birth, male infant, and a multiple delivery. Standard errors in parentheses are clustered at the level of the primary health service area. There are 180 clusters. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Table A.5: Intention-to-treat effect on child weight and length

	Ln (weight)		Ln (height)	
	(1) Low dose	(2) High dose	(3) Low dose	(4) High dose
Physician	0.035	-0.008	0.009	-0.000
	(0.023)	(0.026)	(0.019)	(0.018)
Controls	Yes	Yes	Yes	Yes
Observations	2724	3000	2723	2992
Control group mean	1.704	1.704	4.005	4.005

The dependent variables are the natural logs of child weight at endline in kilograms and recumbent child length at endline. MLP denotes clusters randomly assigned an additional mid-level provider; Physician denotes clusters randomly assigned a physician provider. Measurement FE denote measurer, and month of measurement fixed effects to account for differences in proficiency across the data collectors and improvements over time. In the covariate-adjusted models we control for the woman's age, ethnicity, religion, literacy level (based on whether she could read a simple sentence shown to her in English), pregnancy risk, whether she was offered a cash incentive, decision-making authority in the household (a dummy denoting whether the spouse is the sole decision-maker), and the number of assets owned by the household. We also control for birth characteristics including indicators for a first birth, male infant, and a multiple delivery. Models also control for month of survey. Standard errors in parentheses are clustered at the level of the primary health service area. There are 180 clusters. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

Table A.6: Effect of being observed on consultation length

	Observer was present	Observer was absent
	(1)	(2)
New MLP	0.082*	0.002
	(0.047)	(0.058)
Physician	0.281***	0.171**
	(0.052)	(0.069)
Observations	2383	1214
Dep. variable mean	1.000	1.000
p-value (Physician = MLP)	0.004	0.047

This table examines the length of the consultation when a clinical observer was present vs. not. The dependent variable is the natural log of consultation duration in minutes. Each regression controls for provider age, sex, and years of experience, and the following patient characteristics: age, sex, number of presenting symptoms, illness severity, self-reported health, whether it was a new or follow-up visit, and mode of transportation to the health center. MLP denotes clusters randomly assigned an additional mid-level provider; Physician denotes clusters randomly assigned a physician provider. The omitted comparison group consists of clusters assigned an additional mid-level provider. Standard errors in parentheses are clustered at the level of the primary health service area. There are 180 clusters. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

Table A.7: Specific innovations introduced by new health care providers

Non-physicians	**Physicians**
Reintroduced use of standing orders	Brought a change in the handling of some cases e.g. incomplete abortion
Clerking clients	Brought in new ideas in management of convulsion and labour
He has provided mobility for easy access to the interiors	Carrying out some tests not done previously, and management of cases too e.g severe hypertension
Rearrange process of registration in maternity	Thoroughness in clerking of patients
Community mobilization	New line of treatment in some illness e.g fits
More health talks	Brought a new method of delivery and always encouraged on using antiseptic
Division of labour	Towards diagnosis and laboratory management
Division of labour	Improved post abortion care
Advice on general health maintenance	Improved health talks
Give general advice on any kind of issue or case that comes up	Blood transfusion techniques
Cleaning and sanit[ation] of health center environment	Case management
Advice on cleanliness and hygiene	Patients card
Advice on proper sanitation and cleanliness of the environment	Knowledge sharing with other staff
She provide services on overtime [..] at any time of the day	Proper coordination of the hospital and cleanliness
Advice and encourage to approach patients in good manner and behavior	He advised and adhering to clinical cleanness
Advice on sanitation and cleanliness of the environment	[..] requesting for urinalysis on any cases of high blood pressure, and also advice on use of normal saline in dressing
He explained importance of adhering to clinical advice	He does give idea and information on how and what treatment to give to patients when any kind of case arise
Gives advice on environmental sanitation of the health center	Gives advice on general clinical procedures and maintenance
Advice on proper sanitation and cleanliness of the health center	Advice on health environmental cleanliness
	She brought idea of patient treatment chart
	Advice on proper antenatal visit times [..]
	Washing or dressing of injuries with normal saline
	[..] Drafted procedures in ways of handling any antenatal cases
	Emphasis on urinalysis for any cases of high blood pressure
	Advice on using normal saline in [wound] dressing [..]

Table A.8: Did physicians provide more intensive treatment?

	Cost of delivery		Length of stay	
	(1)	(2)	(3)	(4)
Mid-level provider	125.401	126.291	0.013	0.013
	(117.549)	(107.319)	(0.019)	(0.019)
Physician	174.199	205.441*	0.017	0.016
	(123.004)	(109.142)	(0.020)	(0.020)
Controls	No	Yes	No	Yes
Observations	9496	9494	4520	4520
Control group mean	2049.67	2049.67	0.399	0.399
p-value (Physician = MLP)	0.697	0.489	0.801	0.860

Dependent variables are in the table header. Cost of delivery includes all costs paid by the household towards the delivery excluding transportation costs. Costs are in Naira, the local currency (\$1 $\approx$ 360 Naira at time of study). Length of stay is an indicator denoting at least one night spent in the delivery facility (home births are excluded). MLP denotes clusters randomly assigned an additional mid-level provider; Physician denotes clusters randomly assigned a physician provider. The omitted comparison group consists of clusters assigned an additional mid-level provider. In the covariate-adjusted models we control for the woman's age, ethnicity, religion, literacy level (based on whether she could read a simple sentence shown to her in English), pregnancy risk, whether she was offered a cash incentive, decision-making authority in the household (a dummy denoting whether the spouse is the sole decision-maker), and the number of assets owned by the household. We also control for birth characteristics including indicators for a first birth, male infant, and a multiple delivery. Standard errors in parentheses are clustered at the level of the primary health service area. There are 180 clusters. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.

Table A.9: Was there selection into physician care?

	Mid-level provider	Physician	p-value
A. Mother variables			
Age	24.661	25.146	0.01
Hausa/Fulani ethnicity	0.721	0.844	0.49
Moslem	0.845	0.888	0.38
No formal schooling	0.698	0.683	0.44
Cannot read	0.754	0.721	0.57
Husband makes health-care decisions	0.613	0.677	0.21
Number of prior births	1.886	2.283	0.00
Prior stillbirth or newborn death	0.059	0.076	0.89
Months pregnant at enrollment	4.371	4.307	0.71
Offered conditional incentive	0.656	0.715	0.10
Household assets (out of 11)	1.979	2.104	0.35
Household size	5.711	5.854	0.20
Distance to health center (km)	4.786	4.132	0.94
Sample size	4375	473	
Omnibus test (p-value)			0.37
B. Child variables			
Male infant	0.530	0.574	0.10
Multiple birth	0.024	0.021	0.69
Caesarean delivery	0.006	0.004	0.92
Health card available	1.000	1.000	.
Sample size	4301	469	
Omnibus test (p-value)			0.41

We compare the baseline characteristics of women whose pregnancy and delivery care was provided by a mid-level health provider and women whose care was provided at least in part by a physician. This is based on data from women's health cards. The sample consists of women who used formal care during their pregnancy and delivery, i.e., they attended at least one prenatal visit and gave birth in a health facility. p-values are from a test of difference in group means, and are adjusted for clustering.

Table A.10: Did the incentives have differential effects on health care utilization by health worker assignment?

	Any prenatal care		Number of prenatal visits		3 or more prenatal visits		Facility birth	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mid-level provider	-0.013	-0.013	0.126	0.129	0.002	0.003	0.025	0.023
	(0.022)	(0.022)	(0.110)	(0.104)	(0.029)	(0.028)	(0.022)	(0.021)
Physician	-0.002	0.003	0.066	0.085	0.006	0.011	0.026	0.026
	(0.025)	(0.025)	(0.108)	(0.104)	(0.030)	(0.029)	(0.025)	(0.024)
MLP $\times$ incentive	0.040	0.040	0.193	0.194	0.035	0.035	0.031	0.034
	(0.029)	(0.029)	(0.151)	(0.147)	(0.040)	(0.039)	(0.033)	(0.033)
Physician $\times$ incentive	-0.008	-0.011	-0.043	-0.050	-0.024	-0.026	0.024	0.027
	(0.032)	(0.031)	(0.144)	(0.138)	(0.041)	(0.040)	(0.038)	(0.037)
Incentive		0.066^{***}		0.391^{***}		0.133^{***}		0.126^{***}
		(0.021)		(0.094)		(0.027)		(0.023)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	10586	10586	10586	10586	10586	10586	9410	9410
Control group mean	0.747	0.747	2.563	2.563	0.579	0.579	0.382	0.382
p-value (Physician = MLP)	0.125	0.101	0.136	0.115	0.155	0.134	0.867	0.839

Dependent variables are in the table header. Incentives denote women who were randomly offered a $14 payment conditioned on attending at least three prenatal visits, giving birth in a health facility, and attending at least one postnatal visit. MLP denotes clusters randomly assigned an additional mid-level provider; Physician denotes clusters randomly assigned a physician provider. The omitted comparison group consists of clusters not assigned any additional health providers. In the covariate-adjusted models we control for the woman's age, ethnicity, religion, literacy level (based on whether she could read a simple sentence shown to her in English), pregnancy risk, decision-making authority in the household (a dummy denoting whether the spouse is the sole decision-maker), and the number of assets owned by the household. We also control for birth characteristics including indicators for a first birth, male infant, and a multiple delivery. Standard errors in parentheses are clustered at the level of the primary health service area. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.11: Mean characteristics of health care users by treatment arm

	Control	MLP	Physician	MLP=C	P=C	P=MLP	Joint
A: Mother variables							
Age	24.846	24.796	24.520	0.87	0.10	0.11	0.18
Hausa/Fulani ethnicity	0.708	0.688	0.753	0.29	0.24	0.01	0.04
Moslem	0.824	0.807	0.818	0.01	0.33	0.22	0.02
No formal schooling	0.682	0.684	0.668	0.35	0.55	0.80	0.65
Cannot read	0.713	0.758	0.731	0.31	0.61	0.71	0.59
Husband makes health-care decisions	0.608	0.625	0.643	0.57	0.10	0.26	0.27
Number of prior births	1.894	1.845	1.902	0.32	0.82	0.49	0.59
Prior stillbirth or newborn death	0.065	0.056	0.057	0.25	0.34	0.93	0.48
Months pregnant at enrollment	4.430	4.326	4.300	0.19	0.07	0.53	0.19
Offered conditional incentive	0.616	0.610	0.579	0.84	0.10	0.18	0.21
Household assets (out of 11)	2.124	2.046	2.058	0.92	0.79	0.67	0.91
Household size	5.840	5.764	5.522	0.34	0.21	0.73	0.45
Distance to health center (km)	5.277	5.145	5.344	0.56	0.46	0.87	0.74
Sample size	2105	2264	2230				
Omnibus test (p-value)							0.19
B: Child variables							
Male infant	0.526	0.532	0.542	0.60	0.19	0.44	0.41
Multiple birth	0.022	0.026	0.026	0.56	0.43	0.79	0.73
Caesarean delivery	0.011	0.011	0.006	0.61	0.28	0.09	0.22
Health card available	0.730	0.750	0.759	0.97	0.60	0.65	0.85
Sample size	2046	2194	2149				
Omnibus test (p-value)							0.92

Control (C) denotes clusters not assigned any additional health providers; MLP denotes clusters randomly assigned an additional mid-level provider; Physician (P) denotes clusters randomly assigned a physician. In Panel A each observation is a woman with complete data (baseline and endline); variables are from the baseline survey. In Panel B each observation is a liveborn child; variables are from the endline survey. The figures in Columns 4-6 are p-values from a test of difference in group means. Column 7 is the p-value from a joint test of equality. P-values are adjusted for clustering.

Table A.12: What are the returns to physician human capital?

(a) OLS Estimates

	1-day mortality		7-day mortality	
	(1)	(2)	(3)	(4)
Medical care	-0.007	-0.007	-0.008	-0.008*
	(0.004)	(0.004)	(0.005)	(0.005)
Medical care × physician	-0.009*	-0.009*	-0.009	-0.009
	(0.005)	(0.005)	(0.007)	(0.007)
Controls	No	Yes	No	Yes
Observations	9126	9125	9126	9125
Full Physician effect	-0.016***	-0.016***	-0.017**	-0.017**

(b) IV Estimates

	(1)	(2)	(3)	(4)
Medical care	0.011	0.010	0.009	0.006
	(0.041)	(0.043)	(0.045)	(0.047)
Medical care × physician	-0.031	-0.046	-0.041	-0.059
	(0.032)	(0.032)	(0.040)	(0.039)
Controls	No	Yes	No	Yes
Observations	9126	9125	9126	9125
First-stage F-statistic	13.959	13.085	13.959	13.085
Full Physician effect	-0.020	-0.036	-0.032	-0.053

The dependent variables are shown in the table header. 1-day (7-day) mortality denotes an infant death within 1 day (1 week) of birth. Medical care denotes women who attended at least three prenatal visits or gave birth in a health care facility. Physician-provided care is constructed using information reported in the endline survey. Coefficients are probabilities. Covariate-adjusted models include the same set of controls in Table 4 Column 3. Standard errors in parentheses are clustered at the level of the primary health service area. There are 180 clusters. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.