INSTITUTE for REPLICATION

No. 53
I4R DISCUSSION PAPER SERIES

A Replication of "When a Doctor Falls from the Sky: The Impact of Easing Doctor Supply Constraints on Mortality", Okeke E.N. (2023)

Emma McManus
Joseph Richardson
Vasudha Wattal
Ritchie Woodard

August 2023

INSTITUTE for REPLICATION

I4R DISCUSSION PAPER SERIES

I4R DP No. 53

A Replication of "When a Doctor Falls from the Sky: The Impact of Easing Doctor Supply Constraints on Mortality", Okeke E.N. (2023)

Emma McManus[1], Joseph Richardson[2], Vasudha Wattal[1], Ritchie Woodard[3]

[1]University of Manchester/United Kingdom
[2]University of Lancaster/United Kingdom
[3]University of East Anglia/United Kingdom

AUGUST 2023

E-Mail: joerg.peters@rwi-essen.de
RWI – Leibniz Institute for Economic Research
Hohenzollernstraße 1-3
45128 Essen/Germany
www.i4replication.org

ISSN: 2752-1931

A replication of “When a Doctor Falls from the Sky: The Impact of Easing Doctor Supply Constraints on Mortality”, Okeke E. N. (2023)

(Alphabetical order)
Emma McManus* (emma.mcmanus@manchester.ac.uk, University of Manchester)
Joseph Richardson (j.m.richardson@lancaster.ac.uk, University of Lancaster)
Vasudha Wattal (vasudha.wattal@manchester.ac.uk, University of Manchester)
Ritchie Woodard (r.woodard@uea.ac.uk, University of East Anglia)

*Corresponding author

Funding declaration: The authors received no financial support for this research.
Conflicts of interest: The authors report no conflict of interest.

Abstract

Okeke (2023) evaluates a policy experiment conducted in Nigeria, whereby communities were randomly allocated to receive a new doctor at the local public health center. The performance of these centers was compared to other sites which were allocated either a new midlevel health-care provider, or no additional staff. The study finds that communities assigned a new doctor were associated with a decrease in seven-day infant mortality, such a decrease was not observed in communities assigned a midlevel health-care provider. This suggests that it is the ‘quality’ of the additional doctor driving the effects rather than due to a quantity increase of an additional health worker. The size of the mortality reduction increased with increased exposure to the intervention.

We first conduct a computational reproduction, rerunning the original code and data, finding that the results reported in the original study are reproducible. Second, we test the robustness of the results in several ways, by 1) adapting the existing controls to make the results robust to contamination bias, 2) altering and adding to the control variables included, 3) changing the specification or regression technique used, and 4) testing coding grouping and changing how service use was coded. These changes cause little change to the point estimates, although we find that the original paper’s standard errors were overly conservative, and thus the statistical significance of some results was understated.

1. Introduction

Okeke (2023) evaluates a policy experiment conducted in Nigeria, whereby communities were randomly allocated to receive either a newly qualified doctor, a midlevel health-care provider or no additional staff at their local public health center. A total of 180 primary health service areas (subsequently referred to as sites) were recruited, with 60 of these sites randomly allocated to receive a new doctor, 60 allocated to receive a midlevel health-care provider and 60 to receive no additional staff. Doctors were recruited to sites between February and September 2017 and were deployed for a year.

Okeke (2023) uses data from four sources to evaluate the impact of doctor provision on health outcomes. These are:

1) Data collected on child health outcomes: using data from women's health cards and interviews of mothers conducted approximately three months after birth.
2) Health provider surveys which collected data about the health facility and interviewed the new staff members. This data was collected during two visits to the health center, the first shortly after the new staff member was expected to arrive and then again towards the end of their tenure.
3) Observations of healthcare provision at each of the health centers to determine care quality.
4) Audit data from unscheduled visits to the health centers, to identify if they were open and whether the intervention health-care provider was present. On average, each health center received just over three visits over the intervention period.

Okeke's main analysis estimates the effect of provision of a new doctor (X) on the outcome of seven-day infant mortality (Y) for community populations in Nigeria (P), using a linear probability regression model (M). They also investigate if there is an observed dose-effect for children in utero who were exposed to the intervention for different periods of time, along with exploring the effect on other health outcomes such as: 30-day mortality, deaths in utero, birthweight and child length. All analyses were conducted using STATA 17, on an intention to treat basis.

The main results show that in communities where a doctor was randomly assigned, seven-day mortality decreased by between "0.5 and 0.8 percentage points" in terms of the probability of an infant death within the first week of life, compared to sites that did not receive an additional worker (results described in Table 4, page 604 – without controls: 0.53% (0.0036), with basic controls: 0.69% (0.0036), extended controls: 7.8% (0.0035) and double lasso: 7.7% (0.0034)).

In the present paper, we investigate whether Okeke's (2023) analytical results are computationally reproducible and further test the robustness of their results in the following specification checks: 1) estimating specifications robust to potential contamination bias, 2) altering and adding to the control variables included in the main results, 3) changing the specification or regression technique used, and 4) changing how service use was coded.

Okeke (2023) performed a large number of analyses. As such, we chose to concentrate most of our robustness checks on their main analyses, which reported the impact on seven-day infant mortality. We also chose to change the coding of how service use was determined, as there was some uncertainty surrounding the choice of groupings by the replicating authors.

We successfully reproduced all of the main tables reported in the body of the manuscript and the appendix, using the author's code. No discrepancies were found.

In terms of re-estimating specifications to avoid potential contamination bias, we find that the results remain consistent with those originally reported by Okeke, and in some instances become more statistically significant. The point estimate for seven-day infant mortality with basic controls as reported by Okeke was -0.0069 (se: 0.0036, p-value: 0.056) in the robustness check became -0.0075 (se: 0.0025, p-value: 0.003) due to increases in precision. In other robustness analyses where we altered and added to the control variables, we found there to be no changes to the main conclusions and the results remained similar. When investigating the definition of a high vs low dose, we found the conclusions were not sensitive to the cut-off point chosen. The results of the coding change for service use also remained similar to those reported by Okeke (2023).

All code associated with this replication is included in a GitHub repository (McManus, 2023).

Okeke (2023) was published in *The American Economic Review* complete with an extensive replication package, including analysis do files and datasets (raw and cleaned). As such, the original study author was not contacted in relation to this replication.

2. Reproducibility

We successfully reproduced all of the main tables reported in the body of the manuscript and the appendix, using the author's code. No discrepancies were found. We also did not uncover any coding errors.

3. Replication

We conducted the following robustness checks: 1) estimating specifications robust to potential contamination bias, 2) altering and adding to the control variables included, 3) changing the specification or regression technique used, and 4) changing how service use was coded.

The decision to conduct the first of these robustness checks was taken after reading the paper but prior to observing the manuscript code. The subsequent robustness checks were conceived after the original data and code were examined and executed.

3.1 Robustness to Contamination Bias

The results of this robustness check are shown in Tables 1 to 3.

Several specifications in Okeke (2023) include two treatment indicators and a set of control variables, which means a standard linear regression will not necessarily provide an unbiased estimate of the average treatment effect under treatment effect heterogeneity, with estimates potentially being contaminated by the treatment effects of other interventions (Goldsmith-Pinkham et al, 2022). In such instances, an average treatment effect can be recovered by adding interactions between each treatment dummy and the controls, demeaned by their average value of the covariate within their treatment group. Applying this approach, we recover point estimates of very similar magnitudes to the original analyses (Tables 1, 2, 3, corresponding to Tables 3, 4, and 6 in the original paper), but the estimates of doctors' impact on child mortality increase in statistical significance due to a large reduction in the estimated standard errors[1] (Table 2). The reduced standard errors are likely because the standard errors are clustered at the strata level

[1] We also pick up a statistically significant negative effect of receiving a new MLP on the probability of receiving care from a doctor. However, these effects are so small that they are economically insignificant.

and the interactions between treatment dummies and demeaned controls models the treatment effect heterogeneity between clusters, removing any upward bias in the standard errors from treatment effect heterogeneity between clusters[2].

3.2 Extended controls

The results of this robustness check are shown in Tables 4 to 7.

In the empirical analysis, there are two main control specifications used which are called *basic* and *extended* controls. The *basic* controls are variables that related to the mother/pregnancy and consist of: whether an incentive was offered to the mother, mother age dummies, whether this is her first pregnancy, whether she is ethnically Hausa, level of schooling dummies, whether the mother is autonomous or whether the husband makes health-care decisions, whether the household owns a car/truck, whether her last birth was in a health facility, months pregnant at enrollment, and whether the baby was male. While these are comprehensive given the data available, we consider a few changes. Firstly, it is not clear why the author chooses to use age dummies rather than the continuous age variable in the dataset, so we include this instead. Similarly, rather than just considering if this is the mother's first pregnancy, there is a variable on the number of previous births which we include instead. There is also a categorical ethnicity variable which includes more information than if the mother is Hausa or not. In each of these three cases, there is potentially lost information about the mother which we include as a robustness check.

The *extended* controls related to the facilities of the health center. They consist of the average number of deliveries in the last 6 months, whether the facility does cesareans, whether they do blood transfusions and a categorical variable of cleanliness. Here, there are additional variables relating to the health center which we also include to check robustness. These are: whether the facility has running water, number of beds, estimated travel time to referral hospital, whether the facility is open 24/7, whether it has a laboratory, whether it has a pharmacy, number of workers, whether it has a functioning fan/air con, whether it has no source of electricity, the percentage of essential equipment, and a categorical variable of the general condition of the clinic building. We trial some combinations of these control variables but include them all in reported estimates.

Lastly, there are some variables in the data related to whether pregnancy problems were experienced. There are 14 dummies relating to many different symptoms including swelling, weakness, blurred vision, bleeding, excessive vomiting, etc. While pregnancy problems might be partly influenced by the quality of care, there is certainly a random element which will likely impact health outcomes and the likelihood of seeing a doctor. Indeed, if patients experiencing these symptoms look to be transferred to hospitals with a doctor to improve their level of care, then this might present a downward bias on the estimate of the treatment effect. These pregnancy problem dummies could be included in the *basic* controls as they relate to the mother/pregnancy. However, they arguably better fit the description of *extended* controls, so we include them here. It also ensures that we only make changes to the *basic* controls and only make additions to the *extended* controls for our robustness checks. It is most important that all are included in the full specification (the *extended* controls are always added in addition to the *basic* controls).

In general, the results are robust to these changes in the control specifications. While we see small changes in the coefficients, the qualitative conclusions regarding the direction of the effects

[2] Clustered standard errors are biased upwards when there is treatment effect heterogeneity between the clusters (Abadie et al, 2023).

do not change and, for the most part, the statistical significance associated with the coefficients rarely change. The few exceptions were: the double lasso specifications in Table 5, column (4), and Table 6, column (8), where the original coefficients were significant at the 5% level but the replications were marginally insignificant at the 10% level; Table 6, column (7), where the original coefficient was significant at the 1% level but this dropped to the 5% level in the replication; Table 7, column (1) where the original coefficient was significant at a 5% level but this increased to the 1% level in the replication; and Table 7, column (2), where the original coefficient was statistically insignificant but the replication was significant at the 10% level. Overall, altering and adding to the control specification did not change the conclusions of the paper in any meaningful way.

3.3 Regression specification/modelling change

The results of these robustness checks are shown in Tables 8 to 10 and Figure 1.

Table 8 is a replication of Table 5 of Okeke (2023) where the specification is changed to include an interaction between “dosage” and the treatment (i.e., for how long the patient was pregnant while a doctor was assigned), rather than splitting the sample into high and low dosage. As the specifications are quite different, it did not make sense to include the original estimates here as they are not directly comparable. However, the results in Table 8 inform the recoding that occurs in Table 9 where a direct comparison is possible.

Figure 1 shows a plot of the marginal effect of the treatment at different dosages, based on the interaction term included in Table 8. A clear downward trend is displayed, showing that the treatment of introducing a doctor helped reduce the seven-day mortality and even more effectively when the dosage was high. The cut-off used by Okeke (2023) to classify a high dose in their split sample analysis is 5 months or more. This seems quite arbitrary in some ways. Perhaps it is used because it is the median dosage and/or it is linked to this being the first statistically significantly negative estimate in Figure 1. It is interesting that, despite the trend being clearly downward sloping, the estimates at the higher dosages are not found to be statistically significant, perhaps due to larger confidence intervals (as a result of lower sample sizes at these dosages). We considered splitting into three categories of low (less than 4 months), middle (4-6 months) and high (more than 6 months) dosage – linked to the three trimesters of pregnancy. We found both low and middle doses to be insignificant, so we grouped them in Table 9 to present estimates which are more directly comparable to the originals. The overall result is a change in the cut-off point for the definition of high dose from 5 months or more to 7 months or more.

The results in Table 9 are not too different than the originals. The replicated coefficients in the low dosage specifications become smaller and even negative in columns (2) – (4) but remain statistically insignificant. The replicated coefficients for high dosage are generally also more negative (indicating a larger effect) and remain statistically significant – except at slightly lower levels in columns (7) & (8) – likely due to lower sample sizes. However, overall, the results are robust. The effect of the treatment clearly depends on the dosage. At low dosages, there is little to no measurable effect. At higher dosages, there seems to be a statistically significant effect of reducing seven-day mortality. The purpose of this robustness check is not to suggest that 5 and 6 months should be considered low doses, rather to check that they alone were not driving the effect found for high doses. Although the individual estimates for 7-10 months are statistically insignificant due to wide confidence intervals (Figure 1), the effect when grouping them together as the high dose group is statistically significant. This seems to confirm that it is a sample size issue rather than smaller variances in the middle of the distribution or other such possible

explanation. Where the exact cut-off of low vs high dosage should be is less clear, but the author does not make strong claims about this, and a precise definition does not seem necessary.

The final replication in this section is shown in Table 10 and is a replication of Table A.9 in Okeke (2023). It aimed to analyse whether the treatments affected the weight and height of new-born children. The analysis excluded any children which were not alive at birth. It is possible that poor healthcare could lead to underdevelopment and extremely low heights and weights which cause the child not to survive the birth. In an econometric sense, the data may be truncated. Weight and height data were not available for these excluded observations, so we could not re-run the models while including them. Instead, we run truncated regressions with the lower limit at the minimum of the respective dependent variable which we report in Table 10. This did not change the estimated coefficients in any meaningful way, showing the original coefficients and the general findings to be robust.

3.4 Change to coding classification

The results of these robustness checks are shown in Tables 11 and 12.

In Table A.4, Okeke does a decomposition exercise to show whether the infant mortality reductions within each arm are coming from having received care during pregnancy. The author notes that having received care during pregnancy shows lower seven-day mortality among children in the doctor arm by 0.7 percentage points (3.1% mortality rate in control group compared to 2.4% in the doctor arm). The author originally defines received care as going to any institutional setting. We recode the received care variable to show how these mortality reductions are different when care was received at the health care vs another facility. The reason that this might be crucial is if having a doctor in the village led to patients seeking care in various institutional settings which, while may be better than no care or informal care, might still have variable quality levels.

As a reminder, care is defined as having completed at least 3 antenatal visits or having given birth at a facility. Having received care at the health center (either antenatal visits or given birth) seems to contribute to most of the reductions in seven-day mortality. Mortality is 0.6 percentage points lower among children in a doctor village when their mother received care at the health center (2.7% compared to 2.1%). For instances where care was received at a different facility, the mortality rates are higher and not very different in the doctor arm and control group. However, as the author states in the paper, this is merely illustrative since these subgroups are not random. Moreover, we have no information about characteristics of facilities other than the health center which may contribute to quality of care and/or health outcomes.

Next, the author also demonstrates whether there is a substitution across care settings (Table A.15) and finds that there is a small effect of substitution from care at home to that at the health center, but no evidence of substitution across other settings in terms of where children were born. However, as mentioned elsewhere in the paper, the number of those who received care/gave birth at facilities other than the health center was small. We recode the dependent variable to include all non-facility births in one group (at home and other location) and all other facilities in another group (public hospital, other public, private facility) to test whether the results are robust to this definition. The results reported in Table 12 show evidence of switching from non-facility births to health center is nearly 8%, in line with what the author reports. The overall results are robust to recoding of facility vs non-facility births.

4. Conclusion

We were able to computationally reproduce the results of Okeke (2023). Meanwhile, our robustness checks confirmed, and even strengthen, the results of Okeke's analysis, as the point estimates are affected little by our robustness checks while the precision sometimes improves significantly.

References

Abadie, A., Athey, S., Imbens, G.W. and Wooldridge, J.M., 2023. When should you adjust standard errors for clustering?. *The Quarterly Journal of Economics*, *138*(1), pp.1-35.

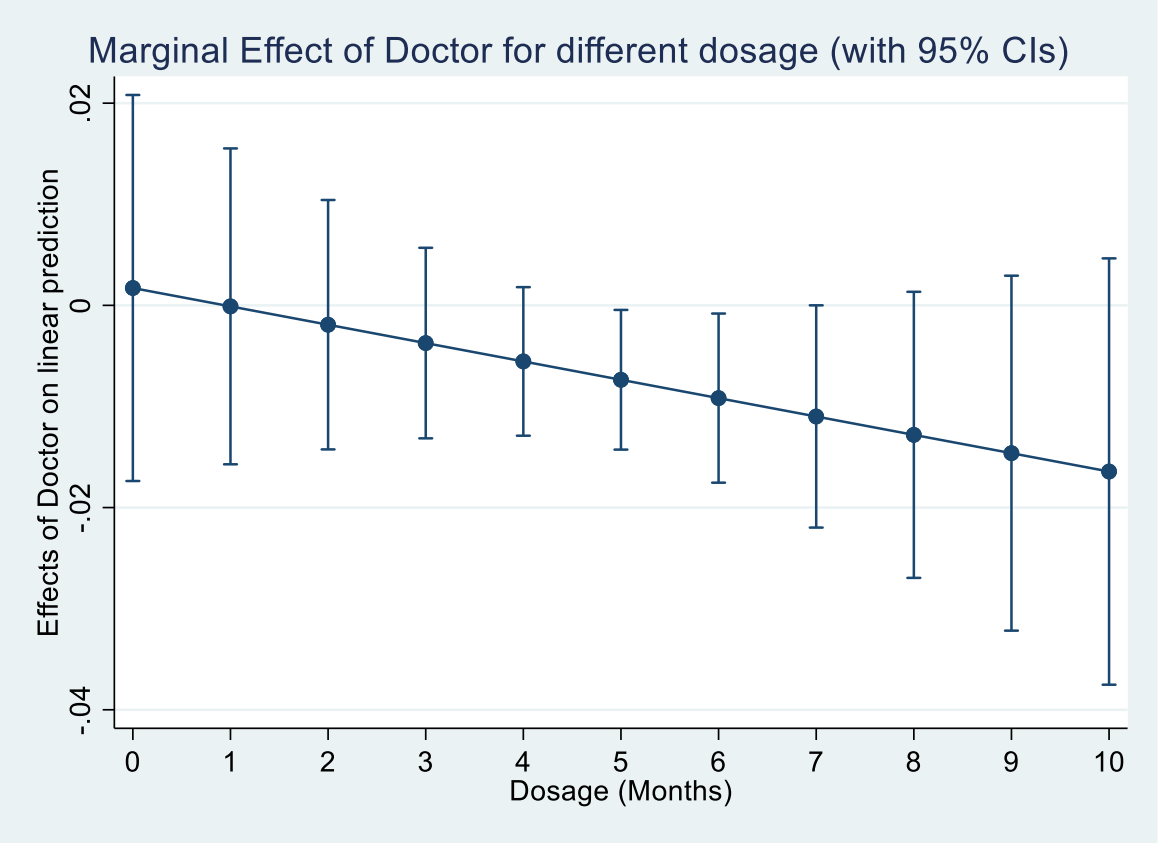
Goldsmith-Pinkham, P., Hull, P. and Kolesár, M., 2022. *Contamination bias in linear regressions* (No. w30108). National Bureau of Economic Research.

McManus, E., 2023. Github repository. Available: https://github.com/e-mcmanus/Okeke23_Replication

Okeke, E.N., 2023. When a Doctor Falls from the Sky: The Impact of Easing Doctor Supply Constraints on Mortality. *American Economic Review*, *113*(3), pp.585-627.

Figures

Figure 1: Marginal effect of Doctor for different 'dosages'



Tables

Table 1: Effect on Seven-Day Mortality, Original analysis (Table 4) compared to Contamination Bias Robust Estimates

	No controls		Basic controls		Extended controls	
	(1)		(2)		(3)	
	Original	**Replicated**	**Original**	**Replicated**	**Original**	**Replicated**
MLP village	−0.0017 (0.0036) {0.636}	-0.0002 (0.0022) {0.914}	−0.0017 (0.0036) {0.855}	0.0003 (0.0022) {0.906}	−0.0005 (0.0037) {0.898}	0.0027 (0.0023) {0.232}
Doctor village	−0.0053 (0.0036) {0.143}	-0.0056** (0.0024) {0.020}	−0.0069* (0.0036) {0.056}	-0.0075*** (0.0025) {0.003}	−0.0078** (0.0035) {0.027}	-0.0086*** (0.0021) {0.000}
Observations	9,126	9,126	9,124	9,125	9,124	9,125
Control group mean	0.0362	0.0362	0.0363	0.0363	0.0363	0.0363

Notes: Original results reported in Table 4, page 604, in Okeke (2023). The replicated results include interactions between the treatment dummies and demeaned controls to prevent contamination bias. We do not provide a lasso specification.
The replicated results for ‘basic controls’ and ‘extended controls’ have a small difference in sample size (9,125 rather than 9,124) as we had to use the ‘reg’ STATA command instead of ‘reghdfe’.
Standard errors are in regular brackets and p-values are in curly brackets, significant at the ***[1%] **[5%] *[10%] level.

Table 2: Effect on Probability that Health Care was Provided by a Doctor, Original analysis (Table 3) compared to Contamination Bias Robust Estimates

	Based on Women's Health Card				Based on Self-Report			
	(1)		(2)		(3)		(4)	
	Original	**Replicated**	**Original**	**Replicated**	**Original**	**Replicated**	**Original**	**Replicated**
MLP village	0.002 (0.016) {0.889}	-0.001** (0.001) {0.028}	0.002 (0.016) {0.892}	-0.002*** (0.001) {0.004}	−0.002 (0.013) {0.859}	-0.003 (0.006) {0.604}	−0.001 (0.013) {0.967}	-0.002 (0.005) {0.705}
Doctor village	0.216*** (0.028) {0.000}	0.190*** (0.011) {0.000}	0.217*** (0.028) {0.000}	0.217*** (0.028) {0.000}	0.083*** (0.016) {0.000}	0.082*** (0.006) {0.000}	0.084*** (0.016) {0.000}	0.084*** (0.006) {0.000}
Controls	No	No	Yes	Yes	No	No	Yes	Yes
Observations	6,891	6,891	6,891	6,891	10,586	10,586	10,586	10,586
Control group mean	0.003	0.003	0.003	0.003	0.058	0.058	0.058	0.058

Notes: These specifications are analogous to Table 3, page 602, in Okeke (2023), except they include interactions between the treatment dummies and demeaned controls to prevent contamination bias.
Standard errors are in regular brackets and p-values are in curly brackets, significant at the ***[1%] **[5%] *[10%] level.

Table 3: Effect on Observed Quality of Treatment, Original analysis (Table 6) compared to Contamination Bias Robust Estimates

	Adherence to fever protocol		Carried out physical exam		Made a diagnosis		Prescribed injection		Prescribed antibiotic		Log of consultation time		Patient communication	
	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
	Original	**Replicated**	**Original**	**Replicated**	**Original**	**Replicated**	**Original**	**Replicated**	**Original**	**Replicated**	**Original**	**Replicated**	**Original**	**Replicated**
MLP village	0.005 (0.023) {0.835}	0.010 (0.018) {0.584}	−0.001 (0.036) {0.984}	0.004 (0.032) {0.892}	−0.068* (0.038) {0.073}	−0.059* (0.035) {0.096}	0.000 (0.031) {0.998}	0.002 (0.030) {0.957}	−0.009 (0.032) {0.789}	−0.012 (0.031) {0.699}	0.092* (0.047) {0.053}	0.093* (0.047) {0.051}	−0.011 (0.019) {0.550}	−0.008 (0.017) {0.604}
Doctor village	0.146*** (0.023) {0.000}	0.167*** (0.020) {0.000}	0.111*** (0.037) {0.004}	0.098*** (0.034) {0.005}	0.244*** (0.034) {0.000}	0.245*** (0.027) {0.000}	−0.084*** (0.027) {0.003}	−0.088*** (0.026) {0.001}	−0.119*** (0.035) {0.001}	−0.118*** (0.031) {0.000}	0.291*** (0.050) {0.000}	0.283*** (0.044) {0.000}	0.057*** (0.020) {0.004}	0.063*** (0.018) {0.001}
Observations	1,168	1,168	2,390	2,390	2,388	2,388	2,390	2,390	2,390	2,390	2,381	2,381	2,388	2,388
Dep. variable mean	0.240	0.240	0.743	0.743	0.670	0.670	0.226	0.226	0.458	0.458	2.096	2.096	0.543	0.543

Notes: Original results reported in Table 6, page 613 in Okeke (2023). The replicated results include interactions between the treatment dummies and demeaned controls to prevent contamination bias.
Standard errors are in regular brackets and p-values are in curly brackets, significant at the ***[1%] **[5%] *[10%] level.

Table 4: Effect on Probability that Health Care was Provided by a Doctor, Original analysis (Table 3) compared to Estimates with Altered and Additional Controls

	Based on Women's Health Card				Based on Self-Report			
	(1)	(2)			(3)	(4)		
	Original	**Original**	**Replicated**	**Replicated**	**Original**	**Original**	**Replicated**	**Replicated**
MLP village	0.002 (0.016) {0.889}	0.002 (0.016) {0.892}	0.002 (0.016) {0.906}	0.001 (0.017) {0.938}	−0.002 (0.013) {0.859}	−0.001 (0.013) {0.967}	0.001 (0.012) {0.952}	-0.001 (0.012) {0.936}
Doctor village	0.216*** (0.028) {0.000}	0.217*** (0.028) {0.000}	0.217*** (0.028) {0.000}	0.216*** (0.025) {0.000}	0.083*** (0.016) {0.000}	0.084*** (0.016) {0.000}	0.085*** (0.016) {0.000}	0.090*** (0.015) {0.000}
(Basic) Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Extended Controls	No	No	No	Yes	No	No	No	Yes
Observations	6,891	6,891	6,891	6,891	10,586	10,586	10,586	10,586
Control group mean	0.003	0.003	0.003	0.003	0.058	0.058	0.058	0.058

Notes: These specifications are analogous to Table 3 in Okeke (2023), except they change three variables in the basic controls (age is in continuous form rather than dummies; ethnicity dummies are included rather than a single dummy for Hausa; and the number of prior births is included rather than a dummy of whether first birth) and the following additional controls are added in the extended controls: dummies for whether various pregnancy problems were experienced and whether the health center had various facilities. As only the control variables included in the specifications change, this only affects columns (2) and (4). Standard errors are in regular brackets and p-values are in curly brackets, significant at the ***[1%] **[5%] *[10%] level.

Table 5: Effect on Seven-Day Mortality, Original analysis (Table 4) compared to Estimates with Altered and Additional Controls

	No Controls	Basic Controls		Extended controls		Double lasso	
	(1)	(2)		(3)		(4)	
	Original	**Original**	**Replicated**	**Original**	**Replicated**	**Original**	**Replicated**
MLP village	-0.0017 (0.0036) {0.636}	-0.0017 (0.0036) {0.639}	-0.0011 (0.0036) {0.768}	−0.0005 (0.0037) {0.898}	−0.0008 (0.0036) {0.820}	-0.0005 (0.0037) {0.886}	-0.0017 (0.0037) {0.635}
Doctor village	-0.0053 (0.0036) {0.143}	-0.0069* (0.0036) {0.056}	-0.0061* (0.0036) {0.091}	-0.0078** (0.0035) {0.027}	-0.0081** (0.0036) {0.025}	-0.0077** (0.0034) {0.026}	-0.0061 (0.0038) {0.110}
(Basic) Controls	No	Yes	Yes	Yes	Yes	No	No
Extended Controls	No	No	No	Yes	Yes	No	No
Observations	9,126	9,124	9,124	9,124	9,124	9,125	9,125
Control group mean	0.0362	0.0363	0.0363	0.0363	0.0363	0.0363	0.0363

Notes: These specifications are analogous to Table 4 in Okeke (2023), except they change three variables in the basic controls (age is in continuous form rather than dummies; ethnicity dummies are included rather than a single dummy for Hausa; and the number of prior births is included rather than a dummy of whether first birth) and the following additional controls are added in the extended controls: dummies for whether various pregnancy problems were experienced and whether the health center had various facilities. As only the control variables included in the specifications change, this only affects columns (2) - (4).
Standard errors are in regular brackets and p-values are in curly brackets, significant at the ***[1%] **[5%] *[10%] level.

Table 6: Effect on Seven-Day Mortality by Treatment Dosage, Original analysis (Table 5) compared to Estimates with Altered and Additional Controls

	Low Dose						
	(1)	(2)		(3)		(4)	
	Original	**Original**	**Replicated**	**Original**	**Replicated**	**Original**	**Replicated**
Doctor village	0.006 (0.006) {0.318}	0.005 (0.006) {0.433}	0.007 (0.006) {0.307}	0.002 (0.007) {0.826}	0.004 (0.007) {0.565}	0.002 (0.009) {0.838}	0.002 (0.008) {0.820}
Basic Controls	No	Yes	Yes	Yes	Yes	No	No
Extended Controls	No	No	No	Yes	Yes	No	No
Observations	2915	2915	2915	2915	2915	2918	2918
	High Dose						
	(5)	(6)		(7)		(8)	
	Original	**Original**	**Replicated**	**Original**	**Replicated**	**Original**	**Replicated**
Doctor village	-0.011** (0.005) {0.031}	-0.013** (0.005) {0.013}	-0.014** (0.005) {0.012}	-0.015*** (0.005) {0.006}	−0.009** (0.004) {0.026}	-0.015** (0.007) {0.030}	-0.011 (0.007) {0.106}
Basic Controls	No	Yes	Yes	Yes	Yes	No	No
Extended Controls	No	No	No	Yes	Yes	No	No
Observations	3200	3200	3200	3200	3200	3201	3201

Notes: These specifications are analogous to Table 5 in Okeke (2023), except they change three variables in the basic controls (age is in continuous form rather than dummies; ethnicity dummies are included rather than a single dummy for Hausa; and the number of prior births is included rather than a dummy of whether first birth) and the following additional controls are added in the extended controls: dummies for whether various pregnancy problems were experienced and whether the health center had various facilities. As only the control variables included in the specifications change, this only affects columns (2) - (4) & (6) - (8).
Standard errors are in regular brackets and p-values are in curly brackets, significant at the ***[1%] **[5%] *[10%] level.

Table 7: Provider Quality and Infant Mortality, Original analysis (Table 8) compared to Estimates with Altered and Additional Controls

	OLS				IV			
	(1)		(2)		(3)		(4)	
	Original	**Replicated**	**Original**	**Replicated**	**Original**	**Replicated**	**Original**	**Replicated**
Proficiency score (%)	-0.0025** (0.0011) {0.018}	-0.0027*** (0.0010) {0.008}			-0.0055** (0.0023) {0.017}	−0.0056** (0.0022) {0.013}		
Standardized proficiency			-0.0035 (0.0024) {0.152}	-0.0040* (0.0023) {0.087}			-0.0155** (0.0067) {0.022}	-0.0161** (0.0067) {0.017}
N	9,124	9,124	9,124	9,124	9,124	9,124	9,124	9,124
First-stage F-statistic					51.1751	54.0115	32.0560	32.6696
Control group mean	0.0363	0.0363	0.0363	0.0363	0.0363	0.0363	0.0363	0.0363

Notes: These specifications are analogous to Table 8 in Okeke (2023), except they change three variables in the basic controls (age is in continuous form rather than dummies; ethnicity dummies are included rather than a single dummy for Hausa; and the number of prior births is included rather than a dummy of whether first birth) and the following additional controls are added in the extended controls: dummies for whether various pregnancy problems were experienced and whether the health center had various facilities. Both basic and extended controls are included in each specification in this table.
Standard errors are in regular brackets and p-values are in curly brackets, significant at the ***[1%] **[5%] *[10%] level.

Table 8: Effect on Seven-Day Mortality by Treatment Dosage, Original analysis (Table 5) splits sample rather than using interaction term

	(1)	(2)	(3)	(4)
Doctor Village	0.009 (0.009) {0.335}	0.007 (0.009) {0.455}	0.002 (0.010) {0.860}	0.004 (0.011) {0.742}
Dosage	0.002 (0.002) {0.308}	0.001 (0.002) {0.591}	0.001 (0.003) {0.776}	0.002 (0.002) {0.372}
Doctor x Dosage	-0.003 (0.002) {0.170}	-0.002 (0.002) {0.208}	-0.002 (0.002) {0.348}	-0.002 (0.002) {0.310}
Basic Controls	No	Yes	Yes	No
Extended Controls	No	No	Yes	No
Observations	6,117	6,117	6,117	6,119

Notes: These specifications are analogous to Table 5 in Okeke (2023), except an interaction between Doctor and Dosage is included rather than splitting the sample into High and Low Dose. The Dosage variable had some negative values which were simply coded as “Low dose” by Okeke. However, negative dosage did not seem logical, so they were recoded to zeros (no dosage) for this analysis.
Standard errors are in regular brackets and p-values are in curly brackets, significant at the ***[1%] **[5%] *[10%] level.

Table 9: Effect on Seven-Day Mortality by Treatment Dosage, Original analysis (Table 5) compared to Changed Cut-Off for High Dosage

	Low Dose – original (< 5 months) vs Replicated (< 7 months)							
	(1)		(2)		(3)		(4)	
	Original	**Replicated**	**Original**	**Replicated**	**Original**	**Replicated**	**Original**	**Replicated**
Doctor village	0.006 (0.006) {0.318}	0.001 (0.004) {0.866}	0.005 (0.006) {0.433}	-0.001 (0.004) {0.861}	0.002 (0.007) {0.826}	−0.003 (0.004) {0.470}	0.002 (0.009) {0.838}	-0.002 (0.006) {0.758}
Basic Controls	No	No	Yes	Yes	Yes	Yes	No	No
Extended Controls	No	No	No	No	Yes	Yes	No	No
Observations	2915	4634	2915	4634	2915	4634	2918	4635
	High Dose – original (≥ 5 months) vs Replicated (≥ 7 months)							
	(5)		(6)		(7)		(8)	
	Original	**Replicated**	**Original**	**Replicated**	**Original**	**Replicated**	**Original**	**Replicated**
Doctor village	-0.011** (0.005) {0.031}	-0.019** (0.009) {0.035}	-0.013** (0.005) {0.013}	-0.021** (0.010) {0.037}	-0.015*** (0.005) {0.006}	−0.020** (0.009) {0.028}	-0.015** (0.007) {0.030}	-0.019* (0.011) {0.099}
Basic Controls	No	No	Yes	Yes	Yes	Yes	No	No
Extended Controls	No	No	No	No	Yes	Yes	No	No
Observations	3200	1481	3200	1481	3200	1481	3201	1484

Notes: These specifications are analogous to Table 5 in Okeke (2023), except the cut-off for the definition between low and high dosage is increased by 1 month to 6 months.
Standard errors are in regular brackets and p-values are in curly brackets, significant at the ***[1%] **[5%] *[10%] level.

Table 10: Effect on child weight and height, Original analysis (Table A.9) compared to Truncated Regression Estimates

	Ln (weight)					
	(1)		(2)		(3)	
	Original	**Replicated**	**Original**	**Replicated**	**Original**	**Replicated**
MLP village	-0.002 (0.020) {0.902}	-0.002 (0.020) {0.915}	-0.003 (0.020) {0.891}	-0.002 (0.020) {0.903}	-0.006 (0.019) {0.769}	-0.005 (0.019) {0.782}
Doctor village	0.007 (0.018) {0.686}	0.007 (0.018) {0.711}	0.005 (0.017) {0.766}	0.005 (0.018) {0.794}	0.001 (0.017) {0.930}	0.001 (0.017) {0.959}
Basic Controls	No	No	Yes	Yes	Yes	Yes
Extended Controls	No	No	No	No	Yes	Yes
Observations	8534	8530	8534	8530	8534	8530
Control Group Mean	1.704	1.704	1.704	1.704	1.704	1.704
	Ln (height)					
	(1)		(2)		(3)	
	Original	**Replicated**	**Original**	**Replicated**	**Original**	**Replicated**
MLP village	-0.024 (0.016) {0.137}	-0.025 (0.016) {0.126}	-0.025 (0.016) {0.123}	-0.026 (0.016) {0.110}	-0.027** (0.016) {0.087}	-0.028** (0.016) {0.076}
Doctor village	-0.017 (0.015) {0.273}	-0.018 (0.016) {0.249}	-0.018 (0.015) {0.233}	-0.019 (0.015) {0.209}	-0.021 (0.015) {0.167}	-0.022 (0.015) {0.143}
Basic Controls	No	No	Yes	Yes	Yes	Yes
Extended Controls	No	No	No	No	Yes	Yes
Observations	8521	8502	8521	8502	8521	8502
Control Group Mean	4.005	4.005	4.005	4.005	4.005	4.005

Notes: These specifications are analogous to Table A.9 in Okeke (2023), except a truncated regression is run with the lower limit of the minimum of the dependent variable. Truncated regression is not compatible with the reghdfe command used in the original analysis, but instead strata and quarter of birth dummies were included which is equivalent.
Standard errors are in regular brackets and p-values are in curly brackets, significant at the ***[1%] **[5%] *[10%] level.

Table 11: 7-day mortality by whether medical care was received and from whom, Original analysis (Table A.4) compared to coding change to received care

		Original		**Replicated**	
		Received medical care		Received care = Yes	
		No	Yes	Care at health centre	Care at other facility
Control	# children	974	2,033	1,717	316
	# deaths within 1st week	46	63	47	16
	Percent	4.7%	3.1%	2.7%	5.1%
MLP village	# children	847	2,178	1,824	354
	# deaths within 1st week	39	65	44	21
	Percent	4.6%	3.0%	2.4%	5.9%
Doctor village	# children	949	2,145	1,895	250
	# deaths within 1st week	43	52	39	13
	Percent	4.5%	2.4%	2.1%	5.2%

Notes: Original results reported in Table A.4, page 18 in Okeke (2023). The first two columns examine mortality by whether medical care was received during pregnancy in each experimental arm, as reported by Okeke. The second two columns reflect a coding change to separate out where care was received- the health center or other facility, with the aim to check if receiving care at health center contributes to greater mortality reductions

Table 12: Is there evidence of changes in substitution patterns? Original analysis (Table A.15) compared to coding change to received care

	Health center	Non facility	Other facility
	(3)	(1) & (6)	(2), (4) & (5)
	Original	**Replicated**	
MLP village	0.030 (0.020) {0.123}	-0.039** (0.018) {0.034}	0.008 (0.009) {0.370}
Doctor village	0.045** (0.020) {0.027}	-0.034* (0.020) {0.085}	-0.011 (0.009) {0.208}
Observations	9,124	9,124	9,124
Control group mean	0.331	0.614	0.055

Notes: Original results reported in Table A.15, page 29 in Okeke (2023). The original table looks at where a study child was born: at home, public hospital, health center, other public, private facility and other location. In this table, we only show the results of 'Health center'. We recode to two groups: non-facility and other facility. All models include the extended set of controls. Standard errors in parentheses are clustered at the level of the primary health service area. There are 180 sites. P-values are in curly brackets, significant at the ***[1%] **[5%] *[10%] level.