A comment on "Does Forest Loss Increase Human Disease? Evidence from Nigeria"*

Dominik Bursy, Jerico Fiestas-Flores

January 16, 2026

Abstract

Berazneva and Byker (2017) research the effect of deforestation on child disease incidence in Nigeria using data from the 2008 and 2013 DHS survey waves. They report that forest loss in the previous year is associated with a higher incidence of malaria, but find no effect on diarrhea or cough. Using the original processed data and extending the panel to include the 2018 DHS survey, we are unable to reproduce the reported malaria results. Consistent with the original study, we also find no systematic relationship between deforestation and the incidence of diarrhea or cough. Our findings further suggest that the original results are sensitive to model specification, and that potential overspecification may contribute to the reported effects.

Keywords: Environmental Degradation, Deforestation, Malaria, Infant Mortality, Children's Health, Nigeria

JEL codes: I12, I15, J13, O13, O15, Q23

*Authors: Bursy: Freie Universität Berlin (FU Berlin) and DIW Berlin. E-mail: dbursy@diw.de, Fiestas-Flores: Leibniz University Hannover. E-mail: fiestas@umwelt.uni-hannover.de. We declare that we received no financial support and have no conflicts of interest. None of the authors have any personal or professional relationship with the original author(s) as defined in the I4R conflict of interest policy.

1 Introduction

Berazneva and Byker (2017) investigate the impact of deforestation on child health in Nigeria, focusing on three illnesses among children under the age of five: fever (malaria), diarrhea, and respiratory infections (cough). The study contributes to the literature on the health consequences of environmental degradation by estimating the causal effect of forest loss on infectious disease incidence using spatial and temporal variation in Nigeria and data from the 2008 and 2013 Demographic and Health Surveys (DHS).

The authors construct a two-year panel at the level of Nigeria's second smallest administrative units, local government areas (LGAs), using reported disease incidence in the two weeks preceding the survey, along with various demographic indicators. Their main explanatory variable, tree loss, is derived from a high-resolution geospatial raster dataset on global forest change based on time-series analysis of Landsat imagery. Forest loss is defined as a transition from a forest to a non-forest state and is measured as a binary indicator. To align the DHS data with the forest loss measures, the authors construct buffer zones of 5 kilometers around rural DHS clusters and 2 kilometers around urban clusters. Within each buffer, they calculate the share of pixels that experienced forest loss in the survey year and in each of the three preceding years.

They present three different estimation and find that the one with the most demographic and spatial controls provides evidence that the first deforestation lag has a positive and significant impact on malaria, with an increase of 1% causing a 2 percentage point increase in malaria incidence. They do not find any relationship between forest loss and diarrhea or cough incidence. These results demonstrate a causal link between forest loss and malaria and provide information on the timing of exposure to malaria relative to tree loss, motivating future work on the longer-term consequences of forest loss in Nigeria and elsewhere on the continent.

In the present paper, we investigate whether Berazneva and Byker's (2017) main results are reproducible and perform additional robustness checks. Although we did our best to follow the replication instructions to process the original data, we are unable to reproduce the results for malaria incidence in magnitude or significance. We also find that there is no relationship between forest loss and diarrhea or cough. We are able to reproduce Table 1 from Berazneva and Byker (2017) using the analysis dataset provided by the authors. However, we are unable to precisely reproduce the construction of this dataset starting from the original raw data sources.

We identify a minor coding issue in which the study uses values for de-normalizing survey weights that differ from current population figures when constructing the panel, potentially due to the use of an outdated or subsequently revised data source.

In particular, the figures used for the population of women of reproductive age (WRA) and for the number of surveyed individuals in this group differ from updated sources in both survey years. Correcting this issue does not significantly affect the main results based on the analysis dataset. Nevertheless, we employ correctly de-normalized weights in our robustness analyses.

We attempt to replicate the original results and conduct two robustness tests to evaluate the sensitivity of the main findings. Specifically, the original study uses the 2008 and 2013 waves of the Nigerian DHS. We (1) reproduce the results using the original data from the Nigerian DHS program (processed data), (2) expand the dataset by including an additional survey wave from 2018 (extended data), and (3) estimate a model with a reduced set of control variables to limit the risk of over-specification.

The results of our replication differ from those reported in the original study in terms of direction, statistical significance, and magnitude. The documentation provided with the replication materials does not allow for an exact reconstruction of the analysis dataset from the original raw data sources. Although we follow the described procedures as closely as possible, differences between the study's analysis dataset and our processed raw data lead to different estimates of the effect of deforestation. In particular, we find that the effect appears in the contemporaneous year rather than in the one-year lag emphasized in the original paper. Moreover, the estimated effects vary in magnitude across lags, with a larger contemporaneous effect that diminishes over time, with the exception of the third lag. This pattern is consistent with a marginal reduction in malaria incidence rather than an increase. A key source of divergence appears to be differences in the spatial distribution of forest loss: while our data show deforestation largely concentrated in southern Nigeria, the study's analysis dataset suggests forest loss distributed more evenly across the country.

When we omit different variables to prevent a model overspecification caused by multicollinearity between spatial and social demographic data, we find that deforestation has no effect in any of the diseases explored, showing the sensitivity of the original model. This is found using the study's analysis data, as well as our processed and extended data.

Finally, although we are unable to reproduce the original results, future replication studies could explore heterogeneity analysis focusing on LGAs with substantial forest cover, as tree cover must be relatively significant to provide protection against diseases (Laporta et al. 2021). In addition, future analyzes could restrict the sample to poorer and/or rural populations to account for the greater vulnerability of these groups and to assess whether deforestation has heterogeneous effects on disease outcomes (Bauhoff and Busch 2020).

2 Computational Reproducibility

Berazneva and Byker (2017) provide a replication package with the final dataset, code, and documentation of the data preparation process. We follow this documentation closely, reconstructing the dataset from the original sources.

We extend the analysis by adding an additional survey wave and an updated version of the forest loss dataset. The replication relies on four main sources: the 2008 and 2013 Nigerian Demographic and Health Surveys (DHS), High-Resolution Global Maps of 21st-Century Forest Cover Change, the DMSP-OLS Nighttime Lights Time Series, and georeferenced soil data.

In what follows, we describe the geolinking of these sources and the construction of the panel at the level of Nigerian local government areas (LGA).

2.1 Data Replication

Demographic and Health Surveys (DHS) The analysis is based on health data from the Nigeria Demographic and Health Surveys (DHS), nationally representative cross-sectional, geo-referenced surveys for 2008 and 2013 (NPC and ICF, 2009, 2014). Access to the data was granted following a brief request. Mapping the variables used in this analysis to the survey was relatively straightforward, although identifying and incorporating control variables presented more of a challenge. We were able to retrieve all relevant characteristics directly from the Children's Recode dataset and did not require any information from the Household Recode. The Children's Recode data were further extended with geographic information on cluster coordinates, allowing for geo-enrichment such as altitude data. The number of observations and mean disease incidence were fairly consistent across survey rounds. For robustness and replication with updated data, we extend the analysis to include the 2018 DHS survey (NPC and ICF, 2019).

Geolinking To link environmental data with health outcomes, we spatially matched DHS survey clusters to multiple geospatial datasets. Each cluster, defined as the centroid of census-based enumeration areas, is subject to random coordinate displacement by DHS to protect confidentiality: rural clusters are displaced 0-5 km (with 1% displaced up to 10 km), and urban clusters 0-2 km, in a random direction and distance (Perez-Haydrich et al. 2013). To account for this, we constructed 5 km buffers for rural clusters and 2 km buffers for urban clusters. Within each buffer, we extracted raster values for soil characteristics, forest loss, and luminosity. Administrative boundaries at the LGA level are taken from the map of Subnational Administrative Boundaries for Nigeria provided by the Humanitarian Data Exchange (HDX), published by the UN Office for the Coordination of Humanitarian Affairs (OCHA) country office in Nigeria. All datasets were reprojected from EPSG:4326 (geographic coordinates, WGS84) to EPSG:3857 (projected coordinates in meters) to ensure accurate distance-based calculations. For each cluster, raster values were averaged, and results were subsequently aggregated by computing means across clusters within each LGA. This procedure provides a consistent framework for enriching DHS microdata with spatially referenced environmental indicators.

Deforestation Data The regressor of interest, tree loss, is drawn from the high-resolution Global Forest Change dataset, which is based on time-series analysis of Landsat imagery at a spatial resolution of one arc-second (approximately 30 meters at the equator), with annual allocation of forest loss (Hansen et al. 2013, version 1.2).

Forest loss is defined as a transition from forest to non-forest (stand-replacement disturbance). The analysis uses the year of gross forest cover loss event dataset, which disaggregates total forest loss to annual scales. Values are encoded as 0 (no loss) or 1-14, corresponding to loss detected in 2001-2014, respectively. This measure captures gross forest loss, so forest gain in the same year is not excluded. In addition, the authors use the tree canopy cover for the year 2000 dataset, defined as canopy closure for vegetation taller than 5 m. Values are encoded as percentages per output grid cell, ranging from 0 to 100.

The forest data are distributed across four large GeoTIFF files organized into tiles. This structure required scripts to identify whether a DHS cluster polygon was located within a single tile or spanned multiple tiles. We extract all pixel values across the relevant tiles and compute the mean proportion of forest loss for each cluster in a given year.

Figure 1 compares the replicated dataset with the dataset used in the original study. We observe notable differences in the spatial distribution of forest loss, with loss primarily concentrated in southern Nigeria and comparatively limited loss in the northern regions. Although we followed the data processing instructions in the original paper as closely as possible, we were not able to fully reproduce the corresponding figure reported in the original study.

To assess robustness, we extend the analysis to include the 2018 DHS survey. To align with this additional survey wave, we update the main regressor using the Global Forest Change 2000-2019 dataset (Hansen et al. 2013, version 1.7).

Soil Data The study also makes use of soil fertility indicators, namely Soil Organic Carbon (SOC), Soil pH (measured in H_2O), and Cation Exchange Capacity (CEC). The underlying data are provided as raster files that map soil properties across Africa at a 250 m resolution and for different depth intervals (Hengl et al. 2015).

For each DHS cluster, we extract soil property values by averaging all raster cell values within a 5 km buffer for rural locations and a 2 km buffer for urban locations. We then compute the mean across all DHS clusters within a given LGA.

Soil values are retrieved at depths of 0-5 cm and 5-15 cm, and subsequently averaged to obtain values for the 0-15 cm layer. We do not observe any statistically significant differences in means or distributions when applying either a two-sided t-test or the Kolmogorov-Smirnov (KS) test. Summary statistics are presented in Table 1.

Luminosity Data Finally, the authors use nighttime lights data as a proxy for economic activity (Chen and Nordhaus 2011). The data, capturing persistent lighting from cities, towns, and other sites, are obtained from the National Oceanic and Atmospheric Administration-National Geophysical Data Center (US Air Force Weather Agency 2009).

Satellite F16 provides coverage for 2005-2008, and F18 for 2009-2013. We use Version 4 of the DMSP-OLS Nighttime Lights Time Series, constructed in smoothed spatial resolution mode at 30 arc-seconds, which corresponds to approximately 1,000 meters (1 km) at the equator.

For each DHS cluster, we retrieve values by averaging all raster cells within the respective buffer zone, and then take the average across clusters within a given LGA. We do not find significant differences in means or distributions using either a two-sided t-test or the Kolmogorov-Smirnov (KS) test. The descriptive statistics are reported in Table 1.

2.2 Discrepancies Between Analysis and Raw Data

For the summary statistics, we followed the authors' methodology by identifying LGAs containing at least one cluster per DHS wave (2008 and 2013) and subsetting the data accordingly. We then merged the variables into a panel dataset for comparison.

The replication dataset differs slightly in the number of observations, likely reflecting differences in the procedures used to link survey clusters to administrative boundaries. Differences between our dataset and that used in the original study primarily arise from variations in spatial linkage procedures and spatial data preprocessing. In addition, the original study retains 40 observations associated with LGAs containing missing values, which we exclude from the replication sample.

Table 1 reports test statistics (two-sided t-tests and Kolmogorov-Smirnov tests) comparing the dependent variables from the original and replication datasets. For survey-based variables, we fail to reject the null hypothesis of equal means and distributions, indicating no significant differences between the datasets. In contrast, for spatial variables, both means and distributions differ significantly, with discrepancies especially pronounced in the forest loss measures (see Figure 1).

Minor differences also arise in the spatial join between DHS cluster locations and the Nigeria LGA shapefile: the original dataset covers 409 LGAs, whereas our cleaned dataset includes 408, reflecting the exclusion of clusters with missing spatial information. Importantly, cluster locations coincide across both datasets, indicating alignment in the underlying DHS coordinates. For reference, Nigeria contains a total of 774 LGAs.

All the differences mentioned above are likely to have an impact in our replication estimates. As we obtain a different distribution of forest loss we might also find different relationships with children diseases. It is possible that the data used in the original paper has gone through multiple improvements over the years, which would highlight the importance of replication studies in order to verify if results hold over time when improved data is available.

2.3 Regression Model

For our analysis, we rely on the same specifications as those used in the paper. We first estimate a balanced pooled OLS for each disease incidence in young children reported by each surveyed cluster (unit level) in lags for forest loss (3), then add fixed effects of LGA and year and the time-varying controlled variables selected by the original paper. See the original study for additional details. We estimate the following equation:

$$\begin{aligned} Y_{itc} = \alpha_0 + \sum_{j=0}^{3} \beta_j loss_{tc}^{j} + \mathbf{X}'_{itc}\gamma + \mathbf{LGA}'_c\pi + \mu' month_{itc} + \mathbf{DHSyear}'_t\theta \\ + \sum_{m=1}^{12} \lambda_m month_{tc}^{m} \times Region_c \times DHSyear_t + \epsilon_{itc}, \end{aligned} \tag{1}$$

2.4 Results

We start by using the newly cleaned raw data (see Section 2.1) and estimate the de-normalized weights using the procedure used in the original paper with updated survey and population values for women in reproductive age (15-49) as we were not able to access the original paper's sources cited in the replication code (see Appendix Section 6.1). Column 1 in Table 3 shows the original study estimates for its most complete specification, and columns 2 and 3 present the results of our robustness tests, which include the use of processed raw data and the addition of an extra year of analysis (2018). Columns 4 to 6 focus on preventing overspecification in the original dataset, as well as in the processed raw and extended data (see Table 4 and Table 5 for the results for diarrhea and cough).

When using the processed raw data (column 2), we find that the point estimate for fever, which is the key variable in the original study, is significant in the current year rather than in the first-year lag. This suggests that deforestation in the current year is associated with a marginal increase in malaria incidence of 2.954 percentage points. When including the additional survey wave from 2018 (column 3), the contemporaneous effect increases and remains positive, while the lagged effects are smaller and generally insignificant.

Since the study includes several variables in their most complete model, we estimate the effect of deforestation by omitting variables that may be correlated with deforestation or with sociodemographic characteristics. When applying this specification to the original analysis data coefficients remain at a similar level of significance, although their magnitude decreases (column 4). Using the same variables with the processed raw data, we find no significant relationship (column 5). For the extended dataset, the results are very similar (column 6), with the contemporaneous effect of deforestation showing a marginal positive effect on malaria incidence.

Our replication results for the relationship between forest loss and malaria, based on the processed raw and extended datasets, differ from those reported in the original study in terms of statistical significance and magnitude. These differences likely reflect a combination of alternative spatial data preprocessing procedures and the inclusion of an additional survey wave from 2018. Within our replication framework, we do not find consistent evidence of a robust association between deforestation and malaria incidence. Similarly, we find no robust relationship between forest loss and the incidence of diarrhea or cough. While the original study reports a substantial effect of forest loss on malaria, characterized by a dynamic pattern consistent with a temporary ecological disturbance described in the tropical medicine literature, we are unable to reproduce this result under our data and methodological assumptions and therefore do not find clear causal evidence supporting this relationship.

2.4.1 Replication with Processed Raw Data As discussed in Section 2.1, the instructions provided by the original study did not allow us to fully reproduce their analysis dataset from the original source. Although we followed their procedures, discrepancies between their analysis data and our processed raw data yield different estimates for the effect of deforestation (see columns 1-2 in Table 3). First, we find that the effect is present in the current year rather than in the first-year lag, as reported in the original paper. Second, our estimates vary in magnitude across lags: the contemporaneous effect is larger and decreases over time (with the exception of the third lag), suggesting that deforestation may contribute to a marginal reduction in malaria fever. The primary source of the discrepancy appears to be the distribution of forest loss in the datasets. While Nigeria exhibits a clear pattern of deforestation concentrated in the south, the original paper's dataset indicate forest loss distributed across the entire country (see Figure 1).

2.4.2 Extending the Time Period In 2018, the DHS conducted a new survey wave in the country, allowing us to expand our sample and examine how including an additional year affects the sign, magnitude, and statistical significance of deforestation on children's disease incidence. We incorporated updated data on forest loss and calculated forest loss lags using the same procedure as the original study (see Section 2.1). Estimating Equation 1 with the extended dataset, we find that the impact of deforestation is significant in the current year and larger than using previous waves, reflecting an increase in disease incidence (see Table 1). Coefficients remain positive across lags, however, including the additional year does not improve the significance of the results for diarrhea or respiratory infections (cough), which remain largely insignificant.

2.4.3 Reduced Specification Our final robustness check consists of revising the variables included in the original paper's final specification (Columns 3, 6, and 9 in Table 3 of the original paper) to ensure that the studies' results, as well as our own, are robust. We exclude interaction variables already present in the model (region $\times$ month $\times$ year) and decide not to use the variables related to soil or luminosity, as they are potentially collinear with deforestation (Angelsen et al. 2014, Burgess et al. 2012). For the demographic variables, we only include those focused on the children's characteristics, the household head's education, and socioeconomic status, avoiding those that could likely be collinear with each other. Table 7 in the Appendix shows the variables included in this robustness check to avoid overspecification and multicollinearity (Clarke 2005).

Table 3 presents results for the analysis data, processed raw data, and extended dataset (columns 4-6) using this new reduced specification. The original paper's final model seems sensitive to the omission of potentially collinear variables (i.e. household characteristics, number of living children), nonetheless, we find that the proposed specification does not result in significant changes on the estimates obtained with the paper's original final model. Results from our robustness checks with the new specification are consistent and less sensitive to variable selection. Using the extended dataset (column 6), the contemporaneous effect of deforestation on malaria is positive and marginally significant at the 10-percent level, with a 1 percent loss of forest cover in the current year associated with a three percentage point increase in malaria incidence. R-squared values remain similar across fever (malaria), diarrhea, and respiratory infections (cough). Overall, we cannot identify a general relationship between deforestation and changes in children's incidence of malaria, diarrhea, or respiratory infections.

3 Conclusion

We replicate the results of Berazneva and Byker (2017) and compare them with those obtained using the processed raw data and an extended dataset that includes an additional survey wave from 2018. We are unable to reproduce the original results for malaria, as our estimates show sensitivity in the current year rather than in the first-year lag, as reported by the authors. When controlling for potential overspecification, we find that deforestation has no significant effect on malaria, indicating that the original model is highly sensitive to changes in specification. We also find no consistent evidence of a relationship between deforestation and increases in the incidence of diarrhea or cough. While the original paper reported a large impact of forest loss on malaria, with a dynamic pattern consistent with a temporary ecological disturbance described in the tropical medicine literature, our replication does not fully confirm this result, and we find no clear causal evidence supporting this link.

4 Figures

(a) Original

(b) Reproduction

Figure 1: Geographic Variation of Forest Loss in Nigeria by LGA (2001-2012)

Panel (a) displays the estimates from Berazneva and Byker (2017) for the average total tree cover loss between 2001 and 2012 across the 409 LGAs included in their analysis sample. Each LGA is shaded according to the average level of tree cover loss observed across the surveyed clusters. LGAs not observed in either the 2008 or 2013 DHS waves are left unshaded. Panel (b) presents our replication, where we estimate for each cluster the share of forest loss between 2001 and 2012 and then compute the average by LGA.

5 Tables

Table 1: Descriptive Statistics Original and Replication Datasets

Variable	Original Dataset			Replication Dataset			Test Statistics	
	Observations	Mean	Std	Observations	Mean	Std	t-stat	KS-stat
Malaria (fever)	41,409	0.144	0.351	41,737	0.144	0.352	0.135	0.000
Diarrhea	41,458	0.106	0.307	41,786	0.106	0.308	0.173	0.000
Respiratory (cough)	41,354	0.110	0.313	41,682	0.110	0.312	-0.111	0.000
Forest loss								
This year	46,654	0.001	0.004	47,039	0.001	0.002	-17.524***	0.047***
1 year ago	46,654	0.001	0.003	47,039	0.001	0.002	-13.245***	0.052***
2 years ago	46,654	0.001	0.003	47,039	0.001	0.002	-16.803***	0.053***
3 years ago	46,654	0.001	0.002	47,039	0.000	0.001	-17.412***	0.061***
Luminosity								
This year	46,654	0.087	1.492	47,039	-0.098	1.193	-20.860***	0.185***
1 year ago	46,654	0.317	1.565	47,039	0.356	1.362	4.049***	0.113***
2 years ago	46,654	-0.392	2.436	47,039	-0.274	1.989	8.143***	0.085***
Soil Organic Carbon	46,654	12.936	7.879	47,039	12.712	7.512	-4.453***	0.036***
Soil pH	46,654	59.801	3.855	47,039	59.876	3.698	3.010***	0.038***
Cation Exchange Capacity	46,654	10.472	4.236	47,039	10.454	3.905	-0.665	0.058***
Tree cover in 2000	46,654	10.820	13.360	47,039	7.883	9.989	-38.089***	0.113***

Notes: This table presents descriptive statistics for the original and replication datasets, restricted to observations included in both the 2008 and 2013 DHS survey waves. The replication dataset was reconstructed from the raw data sources by following the data processing and analysis instructions provided in the paper's replication directory. Differences in distributions are assessed using two-sided t-tests and Kolmogorov-Smirnov tests. Statistical significance is denoted by $p < 0.10$ (*), $p < 0.05$ (**), and $p < 0.01$ (***).

Table 2: Summary Statistics of Child Illnesses by DHS Year

DHSYEAR	Malaria (Fever)			Diarrhea			Respiratory (Cough)		
	Observations	Mean	Std	Observations	Mean	Std	Observations	Mean	Std
2008	16,892	0.158	0.365	16,910	0.099	0.299	16,874	0.120	0.325
2013	22,237	0.135	0.342	22,260	0.109	0.312	22,200	0.105	0.306
2018	18,286	0.239	0.427	18,287	0.128	0.334	18,292	0.169	0.374
Total	57,415	0.175	0.380	57,457	0.112	0.316	57,366	0.129	0.336

Notes: This table presents descriptive statistics for the main dependent variables, malaria (fever), diarrhea, and respiratory illness (cough), using the processed and extended raw dataset. The sample is restricted to observations from the 2008, 2013, and 2018 DHS survey waves.

Table 3: Replication Results for Malaria (Fever)

Malaria (Fever)	Full Specification			Reduced Specification		
	(1)	(2)	(3)	(4)	(5)	(6)
Forest Loss						
This year	0.466	2.954	3.344*	0.422	2.823	3.280*
	(1.027)	(1.621)	(1.642)	(1.003)	(1.605)	(1.627)
	[0.651]	[0.069]	[0.042]	[0.674]	[0.079]	[0.044]
1 year ago	2.042*	1.091	1.042	1.979*	1.331	1.246
	(0.960)	(1.507)	(1.492)	(0.961)	(1.510)	(1.497)
	[0.034]	[0.469]	[0.485]	[0.040]	[0.378]	[0.406]
2 years ago	0.954	-0.117	0.148	1.096	-0.044	0.088
	(0.950)	(2.319)	(2.342)	(0.985)	(2.270)	(2.315)
	[0.315]	[0.960]	[0.950]	[0.266]	[0.984]	[0.970]
3 years ago	-3.434	1.554	1.499	-3.097	0.805	0.858
	(2.758)	(3.701)	(3.696)	(2.628)	(3.666)	(3.662)
	[0.213]	[0.675]	[0.685]	[0.239]	[0.826]	[0.815]
Constant	0.122	0.255**	0.235**	0.136	0.174**	0.166**
	(0.343)	(0.089)	(0.088)	(0.138)	(0.055)	(0.055)
	[0.722]	[0.004]	[0.008]	[0.326]	[0.002]	[0.002]
Observations	40,675	40,459	37,885	40,880	40,620	38,038
R-squared	0.082	0.081	0.082	0.080	0.080	0.081

Notes: This table reports estimation results for fever (malaria) using the full specification from the original study in columns (1)–(3). The dependent variable is reported disease incidence among children under five. Column (1) reproduces the original results, while columns (2) and (3) present estimates using the processed raw data and the extended dataset including the 2018 DHS wave, respectively. Columns (4)–(6) report results from a reduced specification estimated to address potential over-specification. Column (4) uses the original analysis dataset, and columns (5) and (6) use the processed raw data and the extended dataset, respectively. Variables excluded to mitigate over-specification and multi-collinearity are listed in Table 7 in the Appendix. Robust standard errors clustered at the LGA level are reported in parentheses, and p-values are reported in brackets below the standard errors. Statistical significance is denoted by $p < 0.10$ (*), $p < 0.05$ (**), and $p < 0.01$ (***).

Table 4: Replication Results for Diarrhea

Diarrhea	Full Specification			Reduced Specification		
	(1)	(2)	(3)	(4)	(5)	(6)
Forest Loss						
This year	-0.608	-1.182	-1.259	-0.695	-1.381	-1.399
	(0.591)	(1.068)	(1.056)	(0.576)	(1.027)	(1.016)
	[0.304]	[0.268]	[0.233]	[0.227]	[0.179]	[0.169]
1 year ago	0.633	2.829*	2.674*	0.620	2.787*	2.689*
	(0.617)	(1.236)	(1.235)	(0.600)	(1.207)	(1.207)
	[0.305]	[0.022]	[0.031]	[0.301]	[0.021]	[0.026]
2 years ago	0.244	2.258	2.168	0.394	2.775	2.603
	(0.942)	(2.622)	(2.441)	(0.964)	(2.468)	(2.318)
	[0.796]	[0.389]	[0.375]	[0.683]	[0.261]	[0.262]
3 years ago	0.067	1.761	1.784	0.406	1.324	1.456
	(1.131)	(1.539)	(1.516)	(1.062)	(1.556)	(1.537)
	[0.953]	[0.253]	[0.239]	[0.702]	[0.395]	[0.343]
Constant	-0.183	0.157*	0.166*	0.186**	0.003	0.004
	(0.234)	(0.070)	(0.072)	(0.064)	(0.026)	(0.027)
	[0.433]	[0.025]	[0.021]	[0.003]	[0.917]	[0.894]
Observations	40,715	40,508	37,926	40,922	40,669	38,079
R-squared	0.084	0.084	0.085	0.082	0.082	0.084

Notes: This table reports estimation results for diarrhea using the full specification from the original study in columns (1)–(3). The dependent variable is reported disease incidence among children under five. Column (1) reproduces the original results, while columns (2) and (3) present estimates using the processed raw data and the extended dataset including the 2018 DHS wave, respectively. Columns (4)–(6) report results from a reduced specification estimated to address potential over-specification. Column (4) uses the original analysis dataset, and columns (5) and (6) use the processed raw data and the extended dataset, respectively. Variables excluded to mitigate over-specification and multicollinearity are listed in Table 7 in the Appendix. Robust standard errors clustered at the LGA level are reported in parentheses, and p-values are reported in brackets below the standard errors. Statistical significance is denoted by $p < 0.10$ (*), $p < 0.05$ (**), and $p < 0.01$ (***).

Table 5: Replication Results for Respiratory (Cough)

Respiratory (Cough)	Full Specification			Reduced Specification		
	(1)	(2)	(3)	(4)	(5)	(6)
Forest Loss						
This year	-1.161	-0.818	-0.787	-0.992	-0.111	0.019
	(1.123)	(1.760)	(1.800)	(1.089)	(1.703)	(1.744)
	[0.301]	[0.642]	[0.662]	[0.363]	[0.948]	[0.991]
1 year ago	0.425	-0.002	-0.093	0.582	0.613	0.489
	(0.782)	(1.466)	(1.464)	(0.774)	(1.497)	(1.495)
	[0.587]	[0.999]	[0.949]	[0.453]	[0.682]	[0.743]
2 years ago	-1.085	-4.479*	-4.814*	-1.329	-5.565**	-5.992**
	(0.830)	(2.161)	(2.180)	(0.840)	(2.126)	(2.152)
	[0.192]	[0.038]	[0.027]	[0.114]	[0.009]	[0.005]
3 years ago	0.574	3.160	3.545	0.417	2.602	3.064
	(2.271)	(3.261)	(3.274)	(2.157)	(3.237)	(3.253)
	[0.800]	[0.333]	[0.279]	[0.847]	[0.422]	[0.347]
Constant	0.081	0.164*	0.146	0.286***	0.141**	0.136*
	(0.238)	(0.079)	(0.080)	(0.083)	(0.054)	(0.054)
	[0.733]	[0.038]	[0.068]	[0.001]	[0.009]	[0.012]
Observations	40,618	40,407	37,833	40,823	40,568	37,986
R-squared	0.093	0.092	0.093	0.091	0.090	0.091

Notes: This table reports estimation results for respiratory (cough) using the full specification from the original study in columns (1)–(3). The dependent variable is reported disease incidence among children under five. Column (1) reproduces the original results, while columns (2) and (3) present estimates using the processed raw data and the extended dataset including the 2018 DHS wave, respectively. Columns (4)–(6) report results from a reduced specification estimated to address potential over-specification. Column (4) uses the original analysis dataset, and columns (5) and (6) use the processed raw data and the extended dataset, respectively. Variables excluded to mitigate over-specification and multicollinearity are listed in Table 7 in the Appendix. Robust standard errors clustered at the LGA level are reported in parentheses, and p-values are reported in brackets below the standard errors. Statistical significance is denoted by $p < 0.10$ (*), $p < 0.05$ (**), and $p < 0.01$ (***).

References

Angelsen, A., Jagger, P., Babigumira, R., Belcher, B., Hogarth, N. J., Bauch, S., Börner, J., Smith-Hall, C. and Wunder, S.: 2014, Environmental income and rural livelihoods: A global-comparative analysis, *World Development* **64**, S12–S28.

Bauhoff, S. and Busch, J.: 2020, Does deforestation increase malaria prevalence? evidence from satellite data and health surveys, *World Development* **127**, 104734.

Berazneva, J. and Byker, T. S.: 2017, Does Forest Loss Increase Human Disease? Evidence from Nigeria, *American Economic Review* **107**(5), 516–21.

Burgess, R., Hansen, M., Olken, B., Potapov, P. and Sieber, S.: 2012, The political economy of deforestation in the tropics, *The Quarterly Journal of Economics* **127**(4), 1707–1754.

Chen, X. and Nordhaus, W. D.: 2011, Using luminosity data as a proxy for economic statistics, *Proceedings of the National Academy of Sciences* **108**(21), 8589–8594.
URL: *https://www.pnas.org/doi/abs/10.1073/pnas.1017031108*

Clarke, K. A.: 2005, The phantom menace: Omitted variable bias in econometric research, *Conflict Management and Peace Science* **22**(4), 341–352.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O. and Townshend, J. R. G.: 2013, High-resolution global maps of 21st-century forest cover change, *Science* **342**(6160), 850–853.
URL: *https://www.science.org/doi/abs/10.1126/science.1244693*

Hengl, T., Heuvelink, G. B. M., Kempen, B., Leenaars, J. G. B., Walsh, M. G., Shepherd, K. D., Sila, A., MacMillan, R. A., Mendes de Jesus, J., Tamene, L. and Tondoh, J. E.: 2015, Mapping soil properties of africa at 250 m resolution: Random forests significantly improve current predictions, *PLOS ONE* **10**(6), 1–26.
URL: *https://doi.org/10.1371/journal.pone.0125814*

Laporta, G. Z., Ilacqua, R. C., Bergo, E. S., Chaves, L. S. M., Rodovalho, S. R., Moresco, G. G., Figueira, E. A. G., Massad, E., de Oliveira, T. M. P., Bickersmith, S. A., Conn, J. E. and Sallum, M. A. M.: 2021, Malaria transmission in landscapes with varying deforestation levels and timelines in the amazon: a longitudinal spatiotemporal study, *Nature Scientific Reports* **11**, 6447.

National Population Commission - NPC and ICF: 2019, Nigeria demographic and health survey 2018 - final report.
URL: *http://dhsprogram.com/pubs/pdf/FR359/FR359.pdf*

National Population Commission - NPC/Nigeria and ICF International: 2014, Nigeria demographic and health survey 2013.
URL: *http://dhsprogram.com/pubs/pdf/FR293/FR293.pdf*

National Population Commission - NPC/Nigeria and ICF Macro: 2009, Nigeria demographic and health survey 2008.
URL: *http://dhsprogram.com/pubs/pdf/FR222/FR222.pdf*

Perez-Haydrich, C., Warren, J. L., Burgert, C. R. and Emch, M. E.: 2013, Guidelines on the use of dhs gps data, *Technical Report DHS Spatial Analysis Reports No. 8*, ICF International, Calverton, Maryland, USA.
URL: *http://dhsprogram.com/pubs/pdf/SAR8/SAR8.pdf*

United Nations: 2025, Un population division data porta.
URL: *https://population.un.org/dataportal/*

US Air Force Weather Agency: 2009, Version 4 dmsp-ols nighttime lights time series (2009) image and data processing by noaas national geophysical data center, *NOAA, National Geophysical Data Center* .
URL: *https://www.ngdc.noaa.gov/eog/dmsp/downloadV4composites.html*

6 APPENDIX

6.1 Revised De-Normalization of Survey Weights

The original study recommends de-normalizing survey weights when combining multiple waves of DHS data, which is good practice. However, the paper relies on incorrect population figures for women of reproductive age (WRA), defined as women aged 15-49 years, and the links cited as sources for these figures in the replication package are no longer active. Consequently, we obtained population data for WRA from the United Nations and used these values to re-estimate the de-normalizing weights. Specifically, we use UN data for 2008 (35,882,027), 2013 (41,018,918), and 2018 (47,146,386) (United Nations 2025), which differ from the figures reported in the original study. While this discrepancy does not lead to substantially different point estimates, it may result in biased estimators with larger standard errors. Details of the correction are provided in the do-files included our replication repository. We use the equation proposed by the authors:

$$\text{Weight}_t^{\text{new}} = \frac{\text{PopWRA}_t^{\text{weight}}}{100{,}000} \times \frac{\text{PopWRA}_t^{\text{real}}}{\text{PopWRA}_t^{\text{survey}}}$$

6.2 Replication Package Contents and Reproducibility

Table 6: Replication Package Contents and Reproducibility

Replication Package Item	Fully	Partial	No
Raw data provided			✓
Analysis data provided	✓		
Cleaning code provided			✓
Analysis code provided	✓		
Reproducible from raw data		✓	
Reproducible from analysis data	✓		

Notes: This table summarizes the contents of the replication package provided in Berazneva and Byker (2017). The replication package includes a README file listing links to the raw data; however, because several of these links are no longer accessible and the documentation provides limited guidance on data handling, the study is only partially replicable.

6.3 Variables Included in Each Model Specification

Table 7: Variables Included in Each Model Specification

Variable	Description	Full Specification	Reduced Specification	Reason for Exclusion
forest_loss_0	Present deforestation	Yes	Yes	
orest_loss_1-forest_loss_3	Deforestation 1st, 2nd and 3rd lag	Yes	Yes	
i.DHSYEAR	Year fixed effects	Yes	Yes	
i.LGA	LGA fixed effects	Yes	Yes	
i.month#i.region#i.DHSYEAR	Month X region X year fixed effects	Yes	No	Correlated with LGA and Month fixed effects
i.month	Month fixed effects	Yes	No	
i.month#i.region	Month X region fixed effects	No	Yes	
treecover_2000	Tree cover mean in 2000 (%)	Yes	Yes	
cec_ave_pt	Soil CEC 2.5 & 10cm ave	Yes	No	
ph_ave_pt	Soil pH 2.5 & 10cm ave	Yes	No	Correlation with deforestation
occ_ave_pt	Soil organic carbon 2.5 & 10cm ave	Yes	No	
luminosity_chg_0-_2	Change in luminosity coverage	Yes	No	
altitude	Altitude (m)	Yes	Yes	
no_HH_members	# of household members (hv009)	Yes	Yes	
no_kids_under_5	# of children under 5 in the HH (hv014)	Yes	Yes	
time_to_water	Minutes to nearest water source (hv204)	Yes	Yes	
HH_head_edu_years	Years of education by HH head (hv108_01)	Yes	Yes	
head_HH_age	Age of household head in years (hv220)	Yes	No	Not relevant
toilet	1 HH has a flush toilet, 0 if no (hv205)	Yes	No	
firewood	1 if source of fuel is firewood, straw, 0 if other (hv226)	Yes	No	Correlated with rural and poverty
floor	1 if floor made of earth/sand/dung, 0 if other (hv213)	Yes	No	
rural	1 if rural, 0 if urban (hv025)	Yes	Yes	
poorest	1 if in poorest wealth quintile, 0 if richer (hv270)	Yes	Yes	
age	Child's current age(b8)	Yes	Yes	
age_resp	Age of respondent in years (v012)	Yes	No	Not relevant
edu_years	Respondent years of education (v133)	Yes	No	
no_child_total	# of children ever born (v201)	Yes	No	
no_child_living	# of living children (v218)	Yes	No	Correlated with rural and poverty
resp_slept_net	1 if respondent slept under bednet, 0 if no (v461)	Yes	No	
resp_works	1 if respondent works, 0 if no (v714)	Yes	No	
livewith	# of children respondent lives with (v202 + v203)	Yes	Yes	
christian	1 if respondent is Christian, 0 if not (v130)	Yes	Yes	
muslim	1 if respondent is Muslim, 0 if not (v130)	Yes	Yes	
yoruba	1 if respondent ethnicity is Yoruba, 0 if not (v131)	Yes	Yes	
igbo	1 if respondent ethnicity is Igbo, 0 if not (v131)	Yes	Yes	
hausa	1 if respondent ethnicity is Hausa, 0 if not (v131)	Yes	Yes	
married	1 if respondent is married, 0 if otherwise (v501)	Yes	Yes	
pregnant	1 if respondent is pregnant at interview time, 0 if not (v213)	Yes	Yes	

Notes: This table lists the variables excluded from the full specification of Equation 1 to address potential overspecification caused by multicollinearity between spatial and sociodemographic variables. The final column provides the reason for each exclusion.