PROCEEDINGS

Edited By: Giada Adelfio and Antonino Abbruzzo

GRASPA 2023

GRASPA-SIS BIENNAL CONFERENCE

The Researcher Group for Environmental Statistics of The Italian Statistical Society

TIES EUROPEAN REGIONAL MEETING

The International Environmetrics Society

Palermo, 10-11 July, 2023

Dipartimento di Scienze Economiche Aziendali e Statistiche, Università degli Studi di Palermo

Proceedings of the GRASPA 2023 Conference
Palermo, 10-11 July 2023
Edited by: Giada Adelfio and Antonino Abbruzzo
-
Palermo: Università degli Studi di Palermo.

ISBN: 979-12-210-3389-2

Sponsored by:

The effect of deforestation on infant health: a multilevel mediation analysis

Chiara Di Maria

Department of Economics, Business and Statistics, University of Palermo; chiara.dimaria@unipa.it

Abstract. *Humans have been cutting down trees for sustenance since ancient times, but with the advent of the industrial revolution, the deforestation process has reached alarming rates. It is well known that deforestation has detrimental effects on the environment and biodiversity, but effects on human health are understudied. In this paper, we analysed two samples of Nigerian children collected in two years using a multilevel mediation model. The aim is to assess the impact of forest loss on the likelihood of developing certain diseases and whether this effect can be mediated by environmental characteristics like soil fertility. Results provide mixed evidence of the effects of deforestation on human health and call for further investigation.*

Keywords. *Deforestation; Forest loss; Infant diseases; Mediation analysis.*

1 Introduction

Deforestation is a phenomenon which has characterised and fostered human development: humans have cut trees to use their wood for millennia, but the process has seen an acceleration in the last centuries, primarily because of the industrial revolution. From 1700 to the beginning of the 20th century, the deforestation rates almost doubled, moving from 19 to 30 million hectares lost every decade. In 1900, there was a further acceleration till the ’80s, when deforestation reached its peak of 150 million hectares lost in that decade [13]. More recently there was a slight decrease, but according to the UN Food and Agriculture Organization (FAO) Forest Resources Assessment, global deforestation averaged over the years from 2015 to 2020 was 10 million hectares per year.

Forest loss not only affects climate by increasing the amount of greenhouse gases in the atmosphere, but it also induces drought and soil erosion, and destroys habitats of several different species, leading to loss of biodiversity. In addition, the effects of forest loss also extend to the people living in the forests, as they are deprived of a source of food, medicine and other resources, and they are the most exposed to the threats deriving from environmental degradation connected to the reduced tree coverage.

Recently, scholars have started to investigate the association between deforestation and human health. Some authors analysed whether forest loss influences malaria prevalence [1, 8, 9] and [11] provided a systematic literature review. [7] discusses how deforestation may have led to high concentrations of arsenic in drinking water in Khingan, China, posing risks to the health of local residents. [10] and [4] investigated the effects of deforestation on adult mortality in Brazil and infant mortality in Indonesia, respectively.

In this paper, we use the same data analysed by [2] about forest loss in Nigeria and the effects on some infant diseases: malaria, cough and diarrhea. In contrast to [2], we use a multilevel mediation approach to assess the causal effect of deforestation on the probability of developing the aforementioned diseases and evaluate if it is mediated by soil characteristics. In the next section, we provide a description of the data, in Section 3 we present the methodology used for the analysis and the results obtained and finally we discuss our findings.

2 Data description

The data set contains information about a sample of Nigerian children and the area they live in, and it was obtained by merging data from different sources. Information about children's health and demographics are from the 2008 and 2013 waves of Demographic and Health Surveys (DHS) in Nigeria. Data concern children under five and provide information about their health status, i.e. whether they had fever (malaria), cough or diarrhea in the two weeks before the survey date, about their household (number of members, gender, education and marital status of the household head) and house (floor and heating systems of the house, distance from water sources, presence of a flush toilet and bednets). The DHS primary sampling units in Nigeria are clusters within local government areas (LGAs), from which subjects belonging to different households (HH) were sampled. The study is not longitudinal, then subjects are not re-interviewed. The number of clusters in 2008 is 885, 888 in 2013, with approximately 30 thousand children assessed in each year. Children's health status was not assessed by a physician, but is the status reported by their mothers. Data about deforestation were collected from 2000 to 2014 by Earth observation satellites at a spatial resolution of 30 meters, see [5] for details. Forest loss is defined as the share of pixels with annual forest loss in a buffer zone around each DHS cluster with a radius of 5km. Satellite observations were also the source of luminosity data, while information about the soil composition, pH, organic carbon and cation exchange capacity, which can be considered as proxies of soil fertility, was obtained from the International Soil Reference and Information Centre.

3 Statistical analysis

In this section, we describe the approach used in the analysis, introducing the counterfactual mediation framework, and show the results obtained.

3.1 Methods

Mediation analysis is widely used in several fields to investigate whether the effect of an exposure X on a response of interest Y is transmitted by one or more intermediate variables called mediators M. In our setting, forest loss is the exposure, soil characteristics are potential mediators and children's health outcomes are the response variables. Causal mediational effects can be defined within a counterfactual framework [12]. Counterfactuals are statements referring to situations contrary to the actual ones, like "If we had cut CO_2 emissions in 2000, oceans' temperature would be lower" (but CO_2 emissions were not cut in 2000). To understand if oceans' warming is caused by the missing CO_2 reduction one should compare two scenarios: the actual one, without the reduction, and the counterfactual one, where the reduction was enacted in 2000. Assuming there are no unobserved confounders, if the oceans' temperature does not differ over the two scenarios, then the failure to reduce CO_2 emissions is not the cause of oceans' warming. Vice versa, if the two outcomes corresponding to the two scenarios differ, CO_2 is responsible for the increase in oceans' temperature. Denoting by $Y(x)$ the counterfactual outcome one would observe if X were set to x, the average causal effect of X on Y can be defined as

$$\mathbb{E}[Y(x) - Y(x^*)], \quad x \neq x^*.$$

The most common definition of causal mediational effects are the natural direct and indirect effects (NDE and NIE, respectively), defined as follows:

$$NDE = \mathbb{E}[Y(x, M(x)) - Y(x^*, M(x))] \quad (1)$$

$$NIE = \mathbb{E}[Y(x, M(x)) - Y(x, M(x^*))], \quad (2)$$

where $Y(x, M(x^*))$ is the counterfactual value of the outcome if X were set to x and M to the natural value it would assume under $X = x^*$. To identify these effects, i.e. to express them in terms of observed variables, some assumptions about the absence of unobserved confounders are required, see [6, 12].

The sample design of our data is multilevel, since some variables are measured at the children level (level 1), others at the DHS cluster level (level 2). In particular, both the exposure and the potential mediators are cluster-level variables, while the outcomes are individual-level variables. Then, for each potential mediator and outcome, and for each cluster $c = 1, \ldots, C$, year $t = \{2008, 2013\}$ and subject $i = 1, \ldots, n$, we fit the following multilevel mediation models

$$g_M(\mathbb{E}[M_{ct}]) = \beta_{0t} + \boldsymbol{\beta}'_{1t}\mathbf{x}_{ct} + \boldsymbol{\beta}'_{2t}\mathbf{z}_{ct} \quad (3)$$

$$g_Y(\mathbb{E}[Y_{ict}]) = \gamma_{0ct} + \boldsymbol{\gamma}'_{1t}\mathbf{x}_{ct} + \gamma_{2t}M_{ct} + \boldsymbol{\gamma}'_3\mathbf{z}_{ct} + \boldsymbol{\gamma}'_4\mathbf{w}_{ict}, \quad (4)$$

where g_M and g_Y are known link functions for the mediator and the outcome, respectively, $\mathbf{x}_{ct}$ contains forest loss at time $t, t-1$ and $t-2$ for cluster c, $\mathbf{z}_{ct}$ and $\mathbf{w}_{ict}$ are sets of cluster-specific and subject-specific covariates measured in year t, respectively and the $\boldsymbol{\beta}$'s and $\boldsymbol{\gamma}$'s are fixed regression coefficients, except for γ_{0c}, which is a random intercept capturing cluster-dependent effects. Soil organic carbon and cation exchange capability are left-skewed variables, thus we modelled them using a Gamma distribution with log link function, while soil pH is quite symmetric and we fit a linear model. The variables concerning the three infant diseases of interest are binary (presence/absence of related symptoms), thus we used a multilevel logistic model. The mediating role of each soil characteristic was evaluated one at a time.

Natural mediational effects in Equations (1)-(2) can be estimated by simulating counterfactuals based on models (3)-(4) via a Monte-Carlo procedure, as described in [6].

3.2 Results

Results of the mediator and the outcome models are shown in Tables 1 and 2, respectively. We can see that soil pH, organic carbon and cation exchange capability are correlated with each other, and generally they are not associated to forest loss, except for pH which is negatively associated to forest loss two years before the survey date, and cation exchange in 2013, associated to the loss occurred in the same year. Altitude positively influences pH and negatively organic carbon. Treecover is negatively associated to pH and positively associated to organic carbon.

As regards the outcome variables, soil characteristics do not seem to have an explanatory role, except for pH, which negatively affects cough in both survey years and diarrhea in 2008. Living in a rural area increases the children's probability of suffering from the diseases under investigation, but the relationship is significant just in 2008. Age is negatively and significantly associated with all diseases, meaning that older children are less likely to have fever, cough or diarrhea. Being in the poorest economic class seems to decrease the probability of having cough and diarrhea, while a larger number of household members is negatively associated to fever and cough. The coefficients linked to the use of firewood as fuel, the type of floor and the tree cover in the area where the household is do not show a clear pattern. The month and region where the interview took place are included in the model to account for seasonal and spatial patterns, and they result significant for all outcomes in both survey years.

In light of these results, it seems plausible to carry out a mediation analysis to evaluate if the effect of

	pH		Org. carbon		Cation	
	2008	2013	2008	2013	2008	2013
Intercept	**6.211**	**6.199**	**6.492**	**6.522**	**-1.285**	**-0.791**
	(0.024)	(0.026)	(0.217)	(0.219)	(0.261)	(0.267)
loss_{t-2}	**-11.978**	**-4.896**	2.517	2.749	3.788	0.902
	(6.204)	(2.689)	(7.952)	(3.448)	(8.207)	(3.587)
loss_{t-1}	-1.491	0.256	5.372	2.826	-3.026	2.328
	(2.601)	(3.151)	(3.325)	(4.027)	(3.432)	(4.196)
loss_t	3.676	-3.287	8.845	1.030	3.112	**-7.315**
	(3.580)	(1.880)	(4.581)	(2.410)	(4.728)	(2.490)
pH	-	-	**-0.076**	**-0.076**	**0.051**	**0.044**
			(0.004)	(0.004)	(0.004)	(0.004)
org. carbon	**-0.033**	**-0.032**	-	-	**0.041**	**0.041**
	(0.002)	(0.002)			(0.002)	(0.002)
cation	**0.026**	**0.024**	**0.038**	**0.039**	-	-
	(0.002)	(0.002)	(0.002)	(0.002)		
altitude	**0.006**	**0.014**	**-0.078**	**-0.141**	-0.053	**-0.096**
	(0.003)	(0.003)	(0.034)	(0.044)	(0.035)	(0.046)
treecover_t	**-0.009**	**-0.008**	**0.010**	**0.010**	0.002	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table 1: Estimates and standard errors of coefficients estimated for the mediator models. Bold type font denotes significant coefficients at a 0.05 level.

forest loss on cough and diarrhea is mediated by soil pH, since it is the only significant soil characteristic associated with them. Estimates and confidence intervals of the direct and indirect effects, fixing the treatment level to 50 and 100% of forest loss, are shown in Table 3. In 2008, the direct effect of forest loss is significant for both cough and diarrhea, however, for the former its sign stays negative even when the treatment value increases, while for diarrhea the NDE assumes two different signs according to the treatment value. This may denote the fact that the effect of forest loss is 'protective' from diarrhea up to a certain threshold, from which it becomes harmful. The pattern for cough does not change much in 2013, with direct effects always negative and significant, while for diarrhea the direct effect reduces in magnitude and changing treatment value lead to a loss of significance. Indirect effects do not show a clear pattern, and their p-values, not shown in the table, are at the edge of 0.05 significance level.

4 Discussion and conclusions

In this work, we analysed the potential mediating role of soil characteristics in the relationship between forest loss and three diseases very common among children in poor countries like Nigeria. We used a multilevel mediation approach to evaluate the effects of cluster-level environmental variables on individual variables and estimated causal mediational effects. Results are not easy to interpret and sometimes counterintuitive, with forest loss playing in some cases a protective role against diseases. This is however in line with analyses about deforestation effects carried out in other countries. Among the soil characteristics for which information is available, pH seems the one with a relevant association with infant diseases. However, the NIEs are close to the significance level, not allowing us to draw conclusions about the mediating role of pH. These difficulties can be connected to the limitations of the data, such as the non-longitudinal design of the study, which does not allow the evaluation of individual trajectories and a more accurate analysis of causal effects, or the fact that children's diseases are reported by their mothers and not diagnosed by a clinician. These aspects provide suggestions for further directions of research to gain a deeper understanding of the mechanisms underlying forest loss effects on human health.

	Fever		Cough		Diarrhea	
	2008	2013	2008	2013	2008	2013
Intercept	**-2.413**	-1.018	0.545	1.469	**-4.927**	**-2.168**
	(1.002)	(1.035)	(1.039)	(1.217)	(1.118)	(1.059)
ph	0.070	-0.142	**-0.359**	**-0.492**	**0.443**	0.084
	(0.116)	(0.173)	(0.168)	(0.197)	(0.184)	(0.169)
loss_{t-2}	5.467	-7.019	-16.553	-23.855	23.888	3.580
	(25.924)	(11.594)	(8.973)	(13.806)	(30.560)	(11.686)
loss_{t-1}	-13.271	-20.941	-13.946	-23.145	-8.878	-37.053
	(11.424)	(15.745)	(12.471)	(18.230)	(13.520)	(18.723)
loss_{t}	19.053	9.876	22.221	7.587	-9.211	-1.865
	(14.634)	(8.605)	(18.902)	(10.243)	(18.116)	(10.259)
cation	0.004	-0.018	0.018	-0.013	-0.002	-0.017
	(0.010)	(0.010)	(0.010)	(0.012)	(0.011)	(0.011)
org. carbon	-0.017	-0.009	-0.014	0.007	-0.015	-0.006
	(0.009)	(0.010)	(0.010)	(0.011)	(0.012)	(0.011)
rural	**0.378**	0.057	**0.207**	-0.035	**0.204**	-0.028
	(0.079)	(0.078)	(0.082)	(0.092)	(0.092)	(0.082)
time to water	0.003	-0.001	-0.007	0.004	-0.022	0.010
	(0.012)	(0.010)	(0.014)	(0.011)	(0.055)	(0.014)
age	**-0.071**	**-0.120**	**-0.126**	**-0.169**	**-0.203**	**-0.246**
	(0.013)	(0.013)	(0.015)	(0.015)	(0.016)	(0.015)
toilet	-0.036	-0.144	-0.106	-0.044	**-0.318**	**-0.187**
	(0.083)	(0.074)	(0.084)	(0.079)	(0.122)	(0.087)
poorest	-0.085	-0.116	**-0.171**	**-0.166**	**-0.161**	**-0.131**
	(0.059)	(0.062)	(0.070)	(0.076)	(0.066)	(0.064)
# HH members	**-0.012**	**-0.015**	**-0.024**	**-0.014**	-0.007	0.002
	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)
kidnet	-0.083	0.021	-0.030	**0.265**	-0.001	-0.018
	(0.065)	(0.054)	(0.071)	(0.058)	(0.078)	(0.061)
firewood	-0.008	**0.195**	**-0.163**	0.028	**0.181**	**0.229**
	(0.067)	(0.067)	(0.072)	(0.077)	(0.087)	(0.077)
floor	-0.002	0.057	-0.034	**0.151**	0.113	0.078
	(0.054)	(0.052)	(0.061)	(0.060)	(0.064)	(0.057)
altitude	0.009	0.032	-0.008	-0.023	0.022	0.002
	(0.013)	(0.018)	(0.014)	(0.021)	(0.013)	(0.009)
treecover_{t}	0.003	**-0.014**	-0.008	**-0.024**	0.001	-0.007
	(0.004)	(0.005)	(0.004)	(0.005)	(0.006)	(0.005)
month:7	**-0.392**	**-0.245**	**-0.411**	**-0.423**	**-0.497**	**-0.257**
	(0.107)	(0.115)	(0.110)	(0.124)	(0.121)	(0.122)
month:8	**-0.463**	**-0.489**	**-0.570**	**-0.766**	**-0.613**	**-0.540**
	(0.109)	(0.119)	(0.112)	(0.132)	(0.123)	(0.127)
month:9	**-0.588**	**-0.682**	**-0.690**	**-1.047**	**-0.746**	**-0.620**
	(0.115)	(0.118)	(0.119)	(0.131)	(0.131)	(0.126)
month:10	**-0.556**	**-0.894**	**-0.633**	**-1.183**	**-0.802**	**-0.805**
	(0.123)	(0.203)	(0.128)	(0.245)	(0.142)	(0.206)
month:11	-1.823	**-1.658**	-1.367	-2.143	-0.196	-1.686
	(0.997)	(0.706)	(0.885)	(1.169)	(0.988)	(0.687)
region:2	**0.915**	**1.293**	**0.864**	**1.200**	**1.278**	**1.288**
	(0.119)	(0.125)	(0.123)	(0.145)	(0.132)	(0.126)
region:3	**0.520**	**0.291**	**-0.264**	**-0.492**	**0.717**	0.212
	(0.112)	(0.119)	(0.120)	(0.143)	(0.125)	(0.120)
region:4	**1.497**	**1.395**	**1.169**	**0.709**	**0.428**	**0.399**
	(0.140)	(0.156)	(0.145)	(0.177)	(0.177)	(0.163)
region:5	**1.198**	**1.302**	**1.093**	**0.977**	0.143	-0.269
	(0.159)	(0.175)	(0.163)	(0.197)	(0.204)	(0.193)
region:6	0.162	**0.463**	0.216	0.231	**0.612**	0.266
	(0.145)	(0.156)	(0.146)	(0.171)	(0.169)	(0.159)

Table 2: Estimates and standard errors of coefficients estimated for the outcome models. Bold type font denotes significant coefficients at a 0.05 level.

Year	Treat. value	Cough NDE	Cough NIE	Diarrhea NDE	Diarrhea NIE
2008	0.5	**-0.113**	0.033	-0.611	-0.055
		(-0.131, -0.030)	(0.000, 0.420)	(-0.113, 0.899)	(-0.611, 0.000)
	1	**-0.110**	-0.044	**0.853**	-0.109
		(-0.131, -0.052)	(0.000, 0.643)	(0.153, 0.902)	(-0.962, 0.000)
2013	0.5	**-0.147**	0.039	**0.0612**	-0.0766
		(-0.158, -0.132)	(-0.001, 0.183)	(0.059, 0.064)	(-0.131, 0.012)
	1	**-0.135**	0.042	0.669	-0.122
		(-0.164, -0.120)	(-0.349, 0.017)	(-0.111, 0.090)	(-0.922, 0.001)

Table 3: Natural mediational effects of forest loss on cough and diarrhea mediated by pH. Significant effects are written in bold.

References

[1] Bauhoff, S. and Busch, J. (2020). Does deforestation increase malaria prevalence? Evidence from satellite data and health surveys, *World Development*, **127**, 104734.

[2] Berazneva, J., and Byker, T. S. (2017). Does Forest Loss Increase Human Disease? Evidence from Nigeria. *The American Economic Review*, **107**(5), 516–521.

[3] Berazneva, J. and Byker, T. S. (2019). Replication data for: Does Forest Loss Increase Human Disease? Evidence from Nigeria. Nashville, TN: American Economic Association [publisher], 2017. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2019-10-12. https://doi.org/10.3886/E113539V1

[4] Chakrabarti, A. (2021). Deforestation and infant mortality: Evidence from Indonesia, *Economics & Human Biology*, **40**, 100943.

[5] Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A. et al. (2013). High-resolution global maps of 21st-century forest cover change. Science **342**, 850–853.

[6] Imai, K., Keele, L., and Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, **15**(4), 309–334.

[7] Lixin, M., Xuebo, Q., Nan, S. and Guoting, Y. (2014) Human Health Risk of Metals in Drinking-Water Source Areas from a Forest Zone after Long-Term Excessive Deforestation, *Human and Ecological Risk Assessment: An International Journal*, **20**(5), 1200–1212.

[8] MacDonald, A. J. and Mordecai, E. A. (2019). Amazon deforestation drives malaria transmission, and malaria burden reduces forest clearing: a retrospective study, *The Lancet Planetary Health* **3**(1), S13.

[9] Pattanayak, S. K., and Yasuoka, J. (2012). Deforestation and Malaria: Revisiting the Human Ecology Perspective. In *Human Health and Forests* (pp. 219–240). Routledge, London.

[10] Reddington, C., Butt, E., Ridley, D. et al. (2015). Air quality and human health improvements from reductions in deforestation-related fire in Brazil. *Nature Geoscience* **8**, 768–771.

[11] Tucker Lima, J. M., Vittor, A., Rifai, S., and Valle, D. (2017). Does deforestation promote or inhibit malaria transmission in the Amazon? A systematic literature review and critical appraisal of current evidence. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **372**(1722), 20160125.

[12] VanderWeele, T. J. (2015). *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press, Oxford.

[13] Williams, M. (2006). *Deforesting the Earth: From Prehistory to Global Crisis, An Abridgment*, University of Chicago Press, Chicago.