The Journal of Development Studies

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/fjds20

Deforestation and Children: Consequences for Paediatric Malaria and Educational Attainment in Indonesia

Takaaki Kishida, Takahiro Ito & Yuki Yamamoto

To cite this article: Takaaki Kishida, Takahiro Ito & Yuki Yamamoto (25 Sep 2024): Deforestation and Children: Consequences for Paediatric Malaria and Educational Attainment in Indonesia, The Journal of Development Studies, DOI: 10.1080/00220388.2024.2404561

To link to this article: https://doi.org/10.1080/00220388.2024.2404561

The Journal of Development Studies, 2024
Vol. 0, No. 0, 1–18, https://doi.org/10.1080/00220388.2024.2404561

Deforestation and Children: Consequences for Paediatric Malaria and Educational Attainment in Indonesia

TAKAAKI KISHIDA*, TAKAHIRO ITO** & YUKI YAMAMOTO†
*Department of Economics, University of Lausanne, Lausanne, Switzerland; **Graduate School of International Cooperation Studies, Kobe University, Kobe, Japan; †Faculty of Economics, Kansai University, Suita, Japan

(Original version submitted March 2023; final version accepted September 2024)

ABSTRACT *This study examines the effect of forest loss on child health and educational outcomes in Indonesia, a country that has experienced one of the highest deforestation rates in the world. In line with the findings in the medical and biological literature that deforestation can encourage the breeding of malaria-carrying mosquitoes through changes in local ecosystems, our estimation results show that deforestation significantly increases child fever but not other infectious diseases, implying an increased incidence of malaria infection due to deforestation. In addition, the results from the education analysis show that children exposed to larger-scale deforestation in early childhood are afterwards more likely to fall behind academically in terms of years of education. Various robustness checks suggest that these adverse health and educational effects are driven by forest loss but not by other possible preexisting trends or confounders.*

KEYWORDS: Deforestation; child health; education; forest ecosystem; malaria; Indonesia

JEL CLASSIFICATION CODES: Q23; Q56; Q57; I15; I21; O13

1. Introduction

In the past three decades, humans have ravaged 420 million hectares of forest area worldwide through deforestation (equivalent to slightly over 40% of the US's land area), 90 per cent of which has occurred in tropical developing countries (FAO, 2020). Forests are important natural buffers against global warming, and tropical forests are the most species-rich wildlife habitat on the planet (Pimm & Raven, 2000). Therefore, there has been growing concern about climate change associated with rapid deforestation and its detrimental impacts on biodiversity. Meanwhile, the international community has made consistent efforts to address these global environment-related issues as in the various targets in the Millennium Development Goals (MDGs) and the Sustainable Development Goals (SDGs).[1] However, many developing countries, where industrialization is often a top national priority and the enactment of relevant legislation is delayed, face difficulties in preventing environmental destruction.

Correspondence Address: Yuki Yamamoto, Faculty of Economics, Kansai University, 3-3-35 Yamate-cho Suita, Osaka 564-8680, Japan. Email: yamayuki@kansai-u.ac.jp
Supplementary Materials are available for this article which can be accessed via the online version of this journal available at https://doi.org/10.1080/00220388.2024.2404561

In this study, we shed light on another aspect of the consequences of deforestation that has been less addressed—the health and educational effects among children—by exploiting the case of Indonesia, which has one of the world's highest deforestation rates.[2] More specifically, using datasets from nationally representative surveys and satellite-based information on spatiotemporal forest loss, we estimate the impact of forest loss on the incidence of disease (disease symptoms) among children under 5 years of age and on years of education and cognitive ability (measured by test scores) among school-age children.

In the literature, medical and biological studies have found that deforestation has the potential to encourage the breeding of malaria-carrying vectors through changes in local ecosystems and environments (Pattanayak & Pfaff, 2009; Patz, Graczyk, Geller, & Vittor, 2000; Yasuoka & Levins, 2007). In line with these findings, evidence of the linkage between forest loss and malaria *infection* has been reported in recent economics studies. For example, Berazneva and Byker (2017) investigated the effect of forest loss on child malaria in Nigeria and found that the previous year's forest loss increased the incidence of child fever, which may have been caused by malaria. In addition, employing spatial regression techniques, Santos and Almeida (2018) found that deforestation has direct and indirect (spillover) effects on the incidence of malaria in the Brazilian Amazon region.[3] In the case of Indonesia, the country with the second highest number of malaria cases in Asia following India (WHO, 2021), two studies have examined the relationship between forest loss and malaria. Using village-level data on disease outbreaks, Garg (2019) found that the loss of forest cover, measured at the district level, is associated with increased malaria incidence. Chakrabarti (2021) investigated the effect of district-level deforestation on infant mortality and found that firstborn children are at a higher risk of infant mortality than are later-born children when exposed to deforestation-induced increases in malaria in utero.

The principal findings of the present study are as follows. The expansion of deforestation significantly increases child fever, a typical symptom of malaria infection, during the following 12 months. Conversely, falsification tests show that forest loss has no impact on other child diseases, such as cough and diarrhoea. This finding implies that deforestation is unrelated to unobserved trends in child health other than fever, providing evidence of a linkage between deforestation and increased malaria infection. Moreover, the education analysis indicates that early childhood exposure to forest loss lowers subsequent educational attainment in terms of grade levels. Our estimates suggest that when children in an average community are exposed to a 1-standard deviation increase in forest loss during the preschool age period, their malaria incidence increases by 11.1 percentage points and that infected children may subsequently have approximately 0.47 fewer (delayed) years of education.

The contribution of this study to the literature is 2-fold. First, this study adds to the literature by providing evidence on the impact of deforestation using a more rigorous estimation approach. As mentioned above, the link between deforestation and local malaria transmission, and health deterioration has been studied in the Indonesian context (Chakrabarti, 2021; Garg, 2019). Taking these previous studies as a point of departure, this study further attempts to isolate the effects of forest loss from other factors. Specifically, we use tree cover loss measured at an administrative division that is lower than that employed in the two previous studies.[4] The district (*kabupaten*) division used in the previous studies is an administrative unit one level above the subdistrict (*kecamatan*) used in this study, and its average area size is approximately 14 times larger than that of the subdistrict.[5] We can therefore link children to forest data at a finer administrative level. In addition, we eliminate any unobserved heterogeneity across subdistricts within districts by controlling for subdistrict fixed effects, whereas the two studies controlled for district fixed effects. Furthermore, we use the sibling comparison strategy in the education analysis by controlling for sibling fixed effects. Sibling comparison focuses on differences in the timing of exposure within siblings. Therefore, possible heterogeneity between, for example, those who are highly exposed to forest loss (at a given life stage) and those who are not is expected to be small, providing an ideal situation for estimating the educational impact of deforestation.

Second, this study provides new evidence that deforestation is a potential risk factor that threatens not just children's health but their education. Compared to studies examining the impact of deforestation on children's health, studies examining the impact on subsequent education are sparse in the literature.[6] Our estimation results show that early-life exposure to tree cover loss leads to reduced years of education among school-age children, implying that the costs of deforestation may persist into the future. This finding is consistent with a large body of literature documenting that early-life health conditions can have a profound influence on socioeconomic outcomes later in life (Bleakley, 2010a; Currie & Almond, 2011).[7] Thus, the main lesson of this study is that, in addition to the health burden of deforestation, its adverse educational consequences need to be included in the costs of deforestation, which further reinforces the reasonableness of forest conservation policies in terms of social and economic efficiency.

The remainder of this paper proceeds as follows. Section 2 provides background on forest loss and malaria outbreaks based on the Indonesian context and literature. Section 3 outlines the main data used in our analysis, and Section 4 presents the empirical framework. Section 5 discusses the estimation results for the health and education analyses. The conclusion follows in Section 6.

2. Background

2.1. Deforestation and malaria cases in Indonesia

Indonesia had the world's third-largest annual net forest loss (753 thousand ha), behind Brazil (1496 thousand ha) and the Democratic Republic of the Congo (1101 thousand ha), between 2010 and 2020 (FAO, 2020). The loss of forest area in Indonesia has occurred mainly in Sumatra and Kalimantan (Figure 1), with the forest area in these two regions accounting for approximately half of Indonesia's total forest area. The main drivers of Indonesian deforestation in recent decades have been forest clearing for industrial plantations, such as timber, rubber, and oil palm (Austin, Schwantes, Gu, & Kasibhatla, 2019; Wang et al., 2023), and forest fires (Edwards, Naylor, Higgins, & Falcon, 2020). For example, Indonesia's palm oil production in 2019 accounted for approximately 60 per cent of the world's total, a 6-fold increase from 2000 (FAO, 2022). Indonesian deforestation will remain a crucial issue because the demand for palm oil in the international market is expected to reach approximately three to seven times its 2005 level by 2050 (Corley, 2009; Wicke, Sikkema, Dornburg, & Faaij, 2011). In addition, the increase in the number of provinces and other administrative divisions after the fall of the Suharto regime in 1998 is believed to have fostered illegal logging. The decentralization laws took effect in 2001, after Suharto's departure, increasing the number of trees felled and decreasing prices in wood markets due to the increased competition for bribes among local governments (Burgess, Hansen, Olken, Potapov, & Sieber, 2012). Such a deforested environment may take fifty years or more to once again resemble the land as it existed before (Patz et al., 2000).

Malaria transmission in Indonesia has been a serious problem over the past decades, as in other equatorial countries that provide suitable habitat for malaria-carrying mosquitoes. Malaria endemic areas are basically clustered in tropical forest areas, and thus, the number of malaria cases is very limited in Java and the Lesser Sunda Islands, which have negligible forest cover. Around 2010, an estimated 200,000 malaria cases occurred annually, but since then, the number of cases has been gradually decreasing due to the efforts of the international community and government. For example, in 2010, 450 districts reported suspected cases of malaria; in recent years, more than half of them, 266 districts, have eliminated malaria, and many districts have seen a significant decrease in the incidence of malaria transmission (Sitohang et al., 2018).

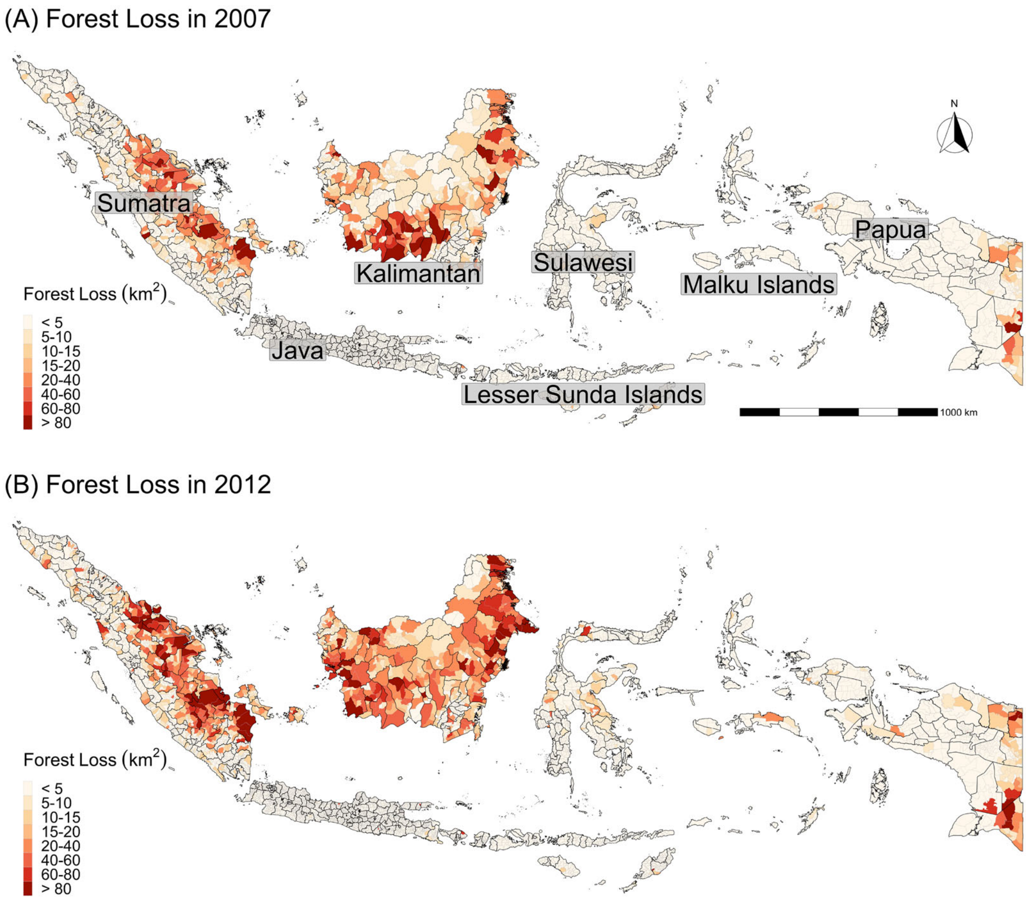


Figure 1. Forest loss in Indonesia.
Notes: (A) shows the forest loss in 2007 and the names of islands, and (B) presents the forest loss in 2012 at the subdistrict level.
Source: Authors' calculations using the forest data constructed by Hansen et al. (2013) and its updated version.

The forest data used in this study mainly cover the years from 2001 to 2012, corresponding to a period with low 'noise' (i.e. regional and political confounders) due to the policy interventions.

2.2. Mechanism of malaria outbreak due to deforestation and its consequences

Mosquito-borne febrile infectious diseases in the tropics include dengue, chikungunya, and Zika, in addition to malaria, the focus of this study. Malaria is transmitted by *Anopheles* mosquito, and the other three are transmitted by *Aedes* species mosquito. While a large body of literature has found that deforestation can facilitate the propagation of malaria-carrying mosquitoes (Yasuoka & Levins, 2007), as mentioned in the introduction, no association between ecological changes caused by deforestation and increased numbers of *Aedes* mosquito larvae has been reported. In the case of dengue fever, for example, two studies found no positive relationship between deforestation in Indonesia and the Brazilian Amazon (Husnina, Archie, Clements, & Wangdi, 2019; Kalbus, de Souza Sampaio, Boenecke, & Reintjes, 2021).

Regarding the mechanisms behind the link between forest clearing and malaria, Pattanayak and Pfaff (2009) and Patz et al. (2000) listed the following four factors. First, after deforestation, the cleared lands are exposed to greater amounts of sunlight and tend to form puddles with a more neutral potential of hydrogen, helping anopheline larvae develop. For example, agriculture that involves standing water, such as paddy cultivation and irrigation, provides new sites suitable for mosquito breeding that are favourable for the development of larvae. Second, increased ground temperatures due to forest loss foster an increased rate of mosquito development into adulthood and an increase in mosquito feeding. Third, the biodiversity loss caused by deforestation can reduce or eliminate the species that prey on Anopheles mosquitoes, thereby indirectly increasing the prevalence of malaria.[8] Fourth, clearing forests is a labour-intensive task and is generally associated with human migration, which can promote malaria transmission. Migrants working in forested areas have lower natural immunity and come into more frequent contact with mosquitoes, and they tend to have reduced access to medical services (Garg, 2019; Pattanayak & Pfaff, 2009).

Malaria is most often transmitted to humans through the bite of a malaria-infected female Anopheles mosquito. The symptoms are mainly fever and flu-like illness. Initially, most people develop symptoms such as fever, sweats, chills, headaches, malaise, muscle aches, and vomiting. Malaria can become a severe, life-threatening disease in a very short time. If it is not treated in time, an infected person may develop kidney failure and mental confusion, fall into a coma, and even die. Treatment relies on access to and administration of prescription drugs, and the time needed to heal varies depending on the person. Furthermore, some empirical studies have documented the effects of early-life exposure to malaria not only on health conditions in adulthood but also on socioeconomic status later in life by reducing literacy, educational attainment, labour productivity, and income (Barreca, 2010; Bleakley, 2010b; Cutler, Fung, Kremer, Singhal, & Vogl, 2010; Lucas, 2010, 2013; Venkataramani, 2012). Despite mounting evidence on the link between malaria and human capital, the impact of childhood exposure to deforestation on subsequent educational attainment remains unaddressed in the literature.

3. Data sources

3.1. Data on child health and educational outcomes

To explore the impact of deforestation on children's health and education, we employ multiple datasets. The main data on child health and educational outcomes are from two rounds of the Indonesian Demographic and Health Survey (DHS) conducted in 2007 and 2012.[9] In these two rounds, respondents' place of residence can be identified only up to the subdistrict (*Kecamatan*) level. Therefore, forest loss is measured at the subdistrict level, as explained in Section 3.2 below. Since the Indonesian DHS is a cross-sectional survey, we construct two-period repeated cross-section data.[10]

In the health analysis, we employ the incidence of disease (disease symptoms) among children under 5 years of age from the two rounds of DHS. Specifically, the occurrence of fever is employed as a proxy for the incidence of malaria since the DHS data contain no information on malaria infection and fever is a typical symptom of malaria. We also use cough and diarrhoea as other child health outcomes. Since these symptoms are unrelated to malaria infection, employing them as dependent variables serves as a falsification test for the linkage between forest loss and child malaria infection. The incidences of fever, cough, and diarrhoea in children are based on mothers' subjective answers to the questions of whether their child suffered from those symptoms during the two weeks preceding the interview. If the mother answers yes, the dependent variable takes on a value of one; otherwise, the value is zero. Thus, the symptoms used in this study are not based on diagnoses of medical professionals and are expected to contain reporting errors. Although we cannot determine the extent to which errors exist, it has been reported that self-reported fever is a good proxy for actual illness (Okiro & Snow, 2010). Moreover, subjective reports by mothers regarding their children's diseases limited to only the two weeks before the survey are likely to underreport actual diseases their children have had, which may lead to underestimation of true impacts.

In the education analysis, we employ years of completed education from the DHS 2012. As explained in Section 3.2 below, tree cover data are available only from 2001. Therefore, the DHS 2007 cannot be used since it does not include school-age children (i.e. 7 years or older) born after 2001, and the sample consists of children born between 2001 and 2005 (children aged 7–11) from the DHS 2012. As a complementary analysis, we also use data from the Indonesian Family Life Survey (IFLS), a nationally representative survey covering 13 of Indonesia's 34 provinces. The IFLS includes data on cognitive test scores, and we examine the influence of forest loss on children's cognitive ability using data from the IFLS-5 (2014–2015). See Appendix A.1 in the Supplementary Material for a detailed description of the IFLS data.

3.2. Forest data

To construct the longitudinal data on forest loss at the subdistrict level, we use the forest data constructed by Hansen et al. (2013), whose updated dataset contains the extent of global tree cover in 2000 and tree cover loss from 2001 to 2020 at a spatial resolution of 30 m.[11] The amount of tree cover in a pixel is defined as a percentage from 0 to 100 per cent based on the definition of a 'tree' as vegetation that is taller than 5 m in height. Tree cover loss in a year is defined as the complete removal of 'trees' from the pixel and is equal to one if it occurred during a given year and zero otherwise. Using these data, we construct the variables for annual tree cover and annual tree cover loss from 2001 to 2014 at the subdistrict level.

Note, however, that it is impossible to calculate the exact area covered by forests or deforested from the data provided by Hansen et al. (2013). For example, suppose that the extent of forest cover in a pixel was 20 per cent in 2000 and that the same pixel became an area with no 'trees' (its forest loss variable takes on a value of one) in 2003. In that case, we cannot determine the exact area of the land that was deforested from 2001 to 2003. We know only that the pixels had no 'trees' in 2003. Nevertheless, considering the relatively fine spatial resolution (30×30 square metres) and the level of human-induced forest loss in Indonesia, deforestation probably did not occur gradually over the years but rather occurred quickly. Therefore, we assume that deforestation occurred in the reported year.[12]

Taking these data limitations into consideration, we construct the forest variables at the subdistrict level as follows. Forest loss in a given year is defined as the sum of pixels that show the complete removal of 'trees' in that year. Thus, tree cover loss in subdistrict j containing M pixels within its border in year t is defined as $Tloss_{jt} = \sum_{m=1}^{M} Loss_{jtm}$, where $Loss_{jtm}$ takes on a value of one if pixel m in subdistrict j lost 'trees' in year t and zero otherwise. Then, the extent of forest cover in year t is calculated as $Tcover_{jt} = \sum_{m=1}^{M} Cover_{jm}^{2000} - \sum_{s=2001}^{t} TLoss_{js}$, where $Cover_{jm}^{2000}$ takes unity if the extent of tree cover in pixel m in subdistrict j was not 0 per cent (i.e. 10% or more) in 2000 and is zero otherwise.

4. Econometric framework

4.1. Empirical specification

In the analysis, we first examine the causal link between forest loss and the incidence of child diseases by estimating the following equation:

$$H_{icjt} = \alpha + \sum_{s=1}^{S} \beta_s \ln Tloss_{js} + \gamma \ln Tcover_{jt} + X_{ijt}\delta + \kappa_j + \lambda_c + \mu_t + u_{icjt}, \quad (1)$$

where H_{icjt} indicates the incidence of disease (fever, cough, or diarrhoea) for child i whose birth cohort is c in subdistrict j and (survey) year-month t. As mentioned in Section 3.1, cough and diarrhoea are used for falsification tests. $\ln Tloss_{js}$, the log of annual tree cover loss, is lagged by S years to allow for changes in the effect of deforestation for a few years after its occurrence.

For example, when $s = 1$, it means the annual forest loss during one year (12 months) preceding the time of survey. In the baseline specification, S is set as 3 to focus on changes in forest loss that have occurred over the past three years, and the sensitivity of our estimates to the choice of lag in the tree cover loss variables is also tested later in the analysis. $\ln Tcover_{jt}$ is the log of tree cover at time t. Since the DHS interview month differs across respondents and between survey rounds,[13] the 12-month period preceding the survey does not correspond exactly to the previous calendar year. Therefore, to adjust for these differences, the forest variables are calculated as a weighted average of the annual forest variables.[14]

X_{ijt} is a vector of child, parent, household, and region characteristics. The list of individual- and household-level controls includes children's age and gender, parents' education, and household wealth index. These variables are potentially important determinants of child health outcomes.[15] For the regional characteristics, we use precipitation at the subdistrict level, population density at the regency/city level, and the GDP and the number of regencies/cities at the province level.[16] In addition, we include subdistrict fixed effects κ_j to control for any unobserved time-invariant characteristics of the subdistricts, such as land area, elevation, and steepness. λ_c and μ_t are the fixed effects for the birth cohort and the year-month of the survey to eliminate time-specific health shocks that are common across all subdistricts at the time of birth or at the time of the survey. u_{ijct} captures the influence of unobserved components.

In analysing the educational attainment of children, letting E_{ihcj} be the educational outcome of child i in household h, birth cohort c, and subdistrict j, we estimate the impact of forest loss in different time periods as follows:

$$E_{ihcj} = \alpha + \sum_{s \in \{\text{pre, sch}\}} \left(\beta_s \ln Tloss_{cjs} + \gamma_s \ln Tcover_{cjs} \right) + X_i\delta + \lambda_h + \mu_c + u_{ihcj}. \tag{2}$$

Unlike the child health analysis using repeated cross-sectional data, the education analysis can be implemented with sibling fixed-effects estimation because of the availability of cohort-varying treatment status. Sibling comparison focuses on differences in the timing of exposure to deforestation within siblings. Therefore, possible heterogeneity between exposed and less exposed children (in their preschool or school years) is expected to be small, providing an ideal situation for estimating the educational impact of deforestation.[17]

The treatment variables, $\ln Tloss_{cjs}$ and $\ln Tcover_{cjs}$, are the log of forest loss and forest cover, respectively, during the period when children of birth cohort c in subdistrict j were preschool-aged ($s =$ pre) or school-aged ($s =$ sch). Note also that we use the *average* annual deforestation area for $\ln Tloss_{cjs}$ and the *average* annual forest cover area for $\ln Tcover_{cjs}$ in each period. Regarding the age categories, we simply focused on the pathways of deforestation impacts before and during schooling. The preschool period, defined here as under seven years of age, is expected to capture the effects of cognitive decline through brain development, and the school period, defined as seven years of age or older, is expected to capture the effects of school absenteeism. However, other age categories can be employed. For example, studies of children's health often focus on the prenatal (in utero) and under-5 periods, and some have found that the period before the age of 3 is the most vulnerable period in terms of the impact of family environment on children's brain development (see, for instance, Dunn et al., 2019). In the analysis, we also check the sensitivity to the age categories. Then, X_i is a vector of individual characteristics such as sex and birth order, μ_c represents the birth-cohort fixed effects, and u_{ihcj} captures unobserved components. The summary statistics for the main empirical variables in the health and education analyses are reported in Table A1 in Supplementary Appendix A.2.

4.2. Identification strategy and issues

To isolate the causal impact of forest loss on health and education, deforestation and its timing need to be exogenous in the sense that they are unrelated to preexisting trends in any health-and education-related outcomes at the individual and subdistrict levels. In the education analysis based on sibling comparisons, this condition seems valid because differences in timing and degree of exposure to forest loss between siblings are considered orthogonal to their potential educational attainment after controlling for their age (birth cohort). Thus, we assume that after controlling for sibling and birth-cohort fixed effects (λ_h and μ_c in Equation 2), forest loss is independent from any observed and unobserved characteristics.

On the other hand, in the health analysis, the above assumption may be overly optimistic because identification relies on subdistrict comparisons. For example, during the study period, deforestation occurred mainly in remote areas. Thus, if deforested areas have a relatively high incidence of infectious disease and a relatively low income level, the subdistricts that experienced extensive deforestation in recent years may exhibit a downwards trend in health. Such heterogeneity in health trends among subdistricts could be captured to some extent by the subdistrict and survey year-month fixed effects (κ_j and μ_τ in Equation 1). However, a more natural expectation is that heterogeneous trends remain to be completely eliminated, causing estimates of the health impact of tree cover loss to be biased.

Thus, given the nonexperimental nature of deforestation, we need to explore its causal impacts more carefully. To check the validity of our identification assumption, we conduct a correlation test (see Supplementary Appendix B.1 for a detailed explanation). In the Supplementary Appendix, we report the partial correlation coefficients between tree cover loss and *ex ante* (i.e. previous year's) child and household characteristics at the subdistrict level controlling for subdistrict fixed effects to eliminate any between-subdistrict heterogeneity. If no correlations are found, deforestation can be assumed to occur independently from preexisting subdistrict-level trends. Table B1 in the Supplementary Appendix shows that the coefficients are quite small and statistically insignificant for all variables, providing evidence that the occurrence of forest loss is not associated with potential trends in child health and educational outcomes or household affluence after eliminating the time-invariant subdistrict heterogeneity.

Furthermore, we verify our identification strategy in three ways. First, we examine the influence of unobserved region-level trends (in Sections 5.1 and 5.2). By estimating Equation (1) with and without several region-level (i.e. subdistrict-, regency/city-, and province-level) controls, we check the robustness of our estimates. If the coefficient estimates for tree cover loss, which is measured at the subdistrict level, change substantially due to the inclusion of the region-level controls, then treatment status (tree cover loss) is correlated not just with the controls but probably with other unobservables. For the region-level controls, we use precipitation at the subdistrict level, population density at the regency/city level, and the GDP and the number of regencies/cities at the province level.[18] Precipitation and population density can be key determinants of infectious diseases. The province-level GDP is used mainly because it is considered that economic development, such as urbanization and industrialization, is associated with deforestation. The number of regencies/cities is to control for political factors related to deforestation since decentralization with the splitting of administrative units was found to be associated with higher deforestation (Burgess et al., 2012).[19] Thus, we indirectly check for potential bias due to heterogeneity in child health and educational outcomes among subdistricts by comparing the estimates from several specifications.

In addition, we further address the possible influences of unobserved regional heterogeneity in child health trends by controlling for regency/city-specific linear trends. Because our data in the health analysis cover only two years and hence *subdistrict*-specific linear trends are collinear with the treatment variables, we control for the linear time trends at the *regency/city* level, calculated as the interaction terms between the survey year and the regency/city dummies. Note that in the education analysis that relies on sibling comparisons with cross-sectional data, we employ the average values of the above region-level variables during the period from birth to the survey date.

Second, as falsification tests, we investigate whether deforestation has effects on diseases other than fever in Section 5.1. If the effect on child fever is truly caused by deforestation-induced malaria infections, then other symptoms unrelated to malaria, such as cough and diarrhoea, should not be affected. Through these falsification tests, we further test the validity of our hypothesis linking tree cover loss to childhood malaria infection.

Finally, we also attempt to address the issue of the selection of unobservables by applying the method proposed by Oster (2019). Even if the coefficients are relatively stable across several specifications, the influence of the remaining selection on unobservables may still be unclear. Under some assumptions, we calculate the bias-adjusted coefficients, in which estimates from the full specification are compared with those from the base specification relying only on the identification assumption, and examine possible biases due to unobserved confounders.

5. Empirical results

5.1. Effects on child health outcomes

We start by exploring the health effects of deforestation. Table 1 shows the estimation results for the effect of tree cover loss on child fever. All specifications include the forest cover in the survey year and subdistrict and year-month fixed effects. In addition, the reported standard errors are clustered at the subdistrict level, which accounts for existing correlations within the same subdistricts. Column (1) presents the estimated impact of tree cover loss over three periods with no additional individual- or region-level controls. The coefficient on tree cover loss during the last 12 months is 0.0415 and is statistically significant at the 5 per cent level, while the coefficients on tree cover loss between one and two years ago and on tree cover loss between two and three years ago are negative and not significant. Notably, these results are not sensitive to the choice of age categories of children or lagged variables for tree cover loss. Using the sample of children that includes newborns and infants aged under seven months or using additional lags of tree cover loss does not substantially change our main results (see Tables B2 and B3 in the Supplementary Appendix).

Table 1. Effect of tree cover loss on the incidence of fever

	Fever		
Dependent variable (sample: children aged 7–59 months)	(1)	(2)	(3)
Tree cover loss (log)			
During the last 12 months	0.0415**	0.0395**	0.0387**
	(0.0180)	(0.0174)	(0.0183)
Between 1 and 2 years ago	−0.0231	−0.0234	−0.0224
	(0.0175)	(0.0171)	(0.0176)
Between 2 and 3 years ago	−0.0164	−0.0161	−0.0160
	(0.0163)	(0.0159)	(0.0159)
Subdistrict and survey year-month FE	Yes	Yes	Yes
Birth-cohort FE	Yes	Yes	Yes
Individual- and household-level controls	No	Yes	Yes
Region-level controls	No	No	Yes
Observations	20,443	20,443	20,443
Number of subdistricts	1540	1540	1540
R-squared	0.018	0.027	0.027

Notes: Standard errors are clustered at the subdistrict level and reported in parentheses. Statistical significance at the 5 per cent level is denoted as **. All columns include the log of the forest cover in the survey year. For brevity, this table presents only the coefficients of interest from Equation (1). ‘Individual- and household-level controls’ include the child age and gender, parental education, and the household wealth index. ‘Region-level controls’ include precipitation at the subdistrict level, population density at the regency/city level, and the number of regencies/cities and GDP at the province level.

We then test the robustness of the coefficients by adding various controls. We control for individual and household characteristics (column 2) and subdistrict-, regency/city-, and province-level variables such as precipitation, population density, GDP, and the number of regencies/cities (column 3). The results show that the estimated coefficients are remarkably stable even when controlling for a rich set of controls, and only tree cover loss during the last 12 months has a positive and significant impact on child fever. Our preferred estimate in column (3) suggests that a 1-standard deviation increase in tree cover loss significantly increases the incidence of fever by 11.1 percentage points.[20]

In addition, the finding that tree cover loss significantly affects the incidence of fever in the short term is consistent with that of previous empirical studies (Berazneva & Byker, 2017; Garg, 2019). A plausible explanation for this short-term impact is that deforested areas may become hostile to mosquitoes in the long term. A loss of biodiversity due to deforestation temporarily reduces the number of mosquito predators, but without the forest environment, mosquitoes themselves cannot survive, and the fever caused by malaria may have consequently decreased. Another probable explanation is that if malaria infections have occurred in an area where forest loss has occurred in the past, people in that area may take various precautions to protect their children from malaria, and as a result, the incidence of malaria may have decreased. Although these pathways cannot be determined in this study, the findings support our assertion that ecological and environmental disturbances due to forest loss inevitably lead to an increase in child malaria infection in the short term.

Then, we conduct falsification tests to investigate whether the health effects of tree cover loss are due to malaria infection. If the increased incidence of fever reported in Table 1 was truly caused by malaria infection, other symptoms unrelated to malaria should not have increased. We examine this possibility using other child health outcomes available in the DHS. Table 2 reports the estimated impacts of tree cover loss on cough and diarrhoea using the same sample and the same set of control variables as in Table 1. Columns (1) to (3) report the results for cough, and columns (4) to (6) report those for diarrhoea. As shown in the table, all the estimated coefficients on tree cover loss during the last 12 months are much smaller in magnitude than those in Table 1, and none of them are significantly different from zero. Thus, the results indicate that deforestation is unrelated to any preexisting trends in child diseases, providing evidence for a link between tree cover loss and increased malaria transmission.[21]

As discussed in Section 4.2, in the health analysis, we further address the possible influence of unobserved regional heterogeneity in child health trends by controlling for regency/city-specific linear time trends. Furthermore, we conduct a sensitivity analysis based on the bias-adjusted coefficient estimation proposed by Oster (2019). The estimation results are reported in Supplementary Appendix B (B.4 and B.5, respectively), indicating that our main results in Table 1 are very robust and not driven by preexisting region-specific trends or possible confounders.

5.2. *Effects on child educational outcomes*

Table 3 shows the estimation results for the effects of tree cover loss on educational outcomes based on Equation (2). All specifications include the (log of) average tree cover during the preschool- and school-age periods with sibling and birth year fixed effects, and the reported standard errors are clustered at the subdistrict level.

Columns (1) to (4) present the impact of tree cover loss on the number of years of completed education for different specifications, using the DHS and IFLS data. Looking at the DHS results (columns 1 and 2), the coefficient estimates for tree cover loss during preschool age are approximately −0.16 to −0.17 and statistically significant at the 10 per cent level. The IFLS results (Columns 3 and 5) also show that the coefficients for tree cover loss during preschool age are negative and statistically significant at the 1 per cent level. Considering the sample size

Table 2. Falsification tests: effect of tree cover loss on the incidence of cough and diarrhoea

	Cough			Diarrhoea		
Dependent variable (sample: children aged 7–59 months)	(1)	(2)	(3)	(4)	(5)	(6)
Tree cover loss (log)						
During the last 12 months	0.0089	0.0077	0.0028	0.0005	−0.0006	0.0030
	(0.0194)	(0.0193)	(0.0206)	(0.0141)	(0.0143)	(0.0144)
Between 1 and 2 years ago	−0.0231	−0.0237	−0.0201	−0.0013	−0.0027	−0.0062
	(0.0194)	(0.0193)	(0.0202)	(0.0119)	(0.0119)	(0.0122)
Between 2 and 3 years ago	−0.0215	−0.0213	−0.0191	−0.0067	−0.0047	−0.0029
	(0.0179)	(0.0178)	(0.0177)	(0.0131)	(0.0133)	(0.0131)
Subdistrict and survey year-month FE	Yes	Yes	Yes	Yes	Yes	Yes
Birth-cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual- and household-level controls	No	Yes	Yes	No	Yes	Yes
Region-level controls	No	No	Yes	No	No	Yes
Observations	20,471	20,471	20,471	20,451	20,451	20,451
Number of subdistricts	1540	1540	1540	1539	1539	1539
R-squared	0.010	0.019	0.022	0.022	0.032	0.032

Notes: Standard errors are clustered at the subdistrict level and reported in parentheses. All columns include the log of the forest cover in the survey year. For brevity, this table presents only the coefficients of interest from Equation (1). 'Individual- and household-level controls' include the child age and gender, parental education, and the household wealth index. 'Region-level controls' include precipitation at the subdistrict level, population density at the regency/city level, and the number of regencies/cities and GDP at the province level.

Table 3. Effect of tree cover loss during preschool and school age on educational attainment

	DHS		IFLS			
	Years of education		Years of education		Cognitive test score	
Dependent variable (sample: children aged 7–14 years)	(1)	(2)	(3)	(4)	(5)	(6)
Tree cover loss (log) during						
Preschool age (6 years or younger)	−0.164*	−0.169*	−0.116***	−0.119***	−0.002	−0.002
	(0.087)	(0.087)	(0.032)	(0.033)	(0.010)	(0.010)
School age	−0.107	−0.111*	−0.023	−0.036	−0.007	−0.009
	(0.066)	(0.065)	(0.022)	(0.022)	(0.008)	(0.007)
Sibling and birth-cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Region-level controls	No	Yes	No	Yes	No	Yes
Observations	2685	2685	775	775	532	532
R-squared	0.903	0.903	0.936	0.938	0.729	0.742

Notes: Standard errors are clustered at the subdistrict level and reported in parentheses. Statistical significance at the 1 per cent level is denoted as *** and at the 10 per cent level as *. For brevity, this table presents only the coefficients of interest from Equation (2). 'Individual controls' include gender and birth order fixed effects and a dummy for missing parent ID. Note that for children with missing parent ID numbers (13 children in six households), we control for household fixed effects instead of sibling fixed effects. Region-level controls include precipitation at the subdistrict level, population density at the regency/city level, and the number of regencies/cities and GDP at the province level.

and representativeness, we expect the DHS results to be more reliable than the IFLS results. The DHS covers the islands of Kalimantan and Sumatra, which have experienced significant deforestation in the recent decades, while the IFLS covers only one of the five provinces in Kalimantan and only about half of the provinces in Sumatra.

The coefficient estimates in the DHS results indicate that a 1-standard deviation increase in tree cover loss in a year during the preschool- and school-age periods reduces the years of schooling by, on average, 0.052 and 0.067 years, respectively.[22] These estimated impacts appear exceedingly small, but this may not necessarily be the case. For example, suppose that children in an average community experience an increase in deforestation by one standard deviation in a year in their school age and that the prevalence of malaria among them increases by 11.1 percentage points (as found in the previous section). In this case, the average decline of 0.052 years then indicates that those additionally infected 11.1 per cent of children have approximately 0.47 fewer years of education.[23]

We further examine the persistency of these adverse educational impacts by including interaction terms between tree cover loss variables and age. However, the result shows that the coefficients of all interaction terms are positive but considerably small and statistically insignificant (not reported here). Although we cannot scrutinize the prolonged impact of deforestation on upper secondary and tertiary education due to data limitations, the result suggests that children who had malaria before school age may enroll in elementary school late, and the one- to two-year delay persists during the sample period without diminishing or enlarging.[24] We also estimate the effects of tree cover loss using a specification that divides the preschool period into early (under 3 years old) and late (3 to 6 years old) periods in Supplementary Appendix B.6. The results in Table B6 in the Supplementary Appendix show no statistical evidence that the effects of deforestation differ between the early and late preschool years.

We then attempt to explore the causal mechanism of the deteriorated impacts of deforestation on education by focusing on influences through cognitive ability (columns 5 and 6 of Table 3). Looking at the impacts of tree cover loss during preschool age, the estimation results show no statistically significant impact on cognitive test scores (the sum of shape matching and math test scores). Although the IFLS has limitations mentioned above, these results indicate that the reduction in educational years found in columns (3) and (4) is not driven by impaired cognitive development from malaria infection.[25] We also estimate the effects of tree cover loss by dividing cognitive test scores into those for shape matching and mathematics, but we found no effects for either (see Table B7 in Supplementary Appendix B.7).

Moreover, we examine the possible influence of the reduced sample size due to our estimation strategy. As mentioned in Sections 4.1 and 4.2, we lost more than half of the observations due to the sibling comparison strategy, which excludes children with no siblings in the sample. We assume that the presence of eligible age siblings is 'as if' random, but if it is related to deforestation, it may create another problem for internal validity, that is, sample selection bias. To test this possibility, we implement subdistrict fixed-effects (not sibling fixed-effects) estimation using the full sample of children (Table B8 in Supplementary Appendix B.8). A detailed discussion of the estimation results is provided in the Supplementary Appendix, but in summary, the results there show that the main results in Table 3 are not driven by limiting the sample.

Finally, we examine the validity of our identification assumption by estimating the bias-adjusted coefficients proposed by Oster (2019). The results are presented in Panel B of Figure B1 in Supplementary Appendix B.5 and show that the DHS estimates in Table 3 are relatively stable across all the conditions examined regarding the influence of unobservables.

6. Conclusion

In this study, we examined the effect of tree cover loss on child malaria infection and subsequent educational attainment using datasets obtained from nationally representative surveys and satellite imagery. Our empirical findings imply that forest loss increased the incidence of child malaria infection in the immediate 12 months and that exposure to deforestation during the preschool-age period subsequently led to fewer years of completed education.

These results have important policy implications. First, malaria outbreaks can be triggered by recent deforestation. Therefore, prompt countermeasures need to be taken in areas with ongoing deforestation or that face the risk of potential deforestation. One effective measure would be the distribution of free mosquito nets to households (Cohen & Dupas, 2010). Second, in addition to the health burden caused by deforestation, its adverse educational consequences need to be included in the costs of deforestation. Although our analysis focuses on children in primary and secondary education due to data limitations, these educational costs may increase in the future because of the accumulation of the negative effect on the attainment of higher levels of education. Understating the cost of deforestation may result in delayed or poorly implemented countermeasures, and therefore, our findings emphasize the importance of adopting effective measures. Furthermore, given the increasing global demand for palm oil products, forest conservation continues to be a challenge in Indonesia and other equatorial countries with expanding oil palm plantations (Corley, 2009; Wicke et al., 2011).

Finally, to guide the direction of future research on estimating the costs of deforestation, we discuss several issues that remain unaddressed in this study. First, the child health outcomes used in the health analysis are drawn from subjective data based on mothers' reports of their children's diseases during the two weeks preceding the survey rather than actual diagnostic data. This is because the DHS data used in the analysis contain no information on the results of a medical examination for malaria parasites. Together with the finding of Okiro and Snow (2010) that a relatively large number of children infected with malaria plasmodium did not have fever on the day of the survey, the health cost of forest loss found in this paper may be underestimated. Second, the education analysis in this study employed a sample of school-aged children due to the limited forest data available, and therefore, the analysis is silent on whether and how long the adverse impacts of tree cover loss persist in secondary and tertiary education or in the labour market. The literature lacks studies examining both the health and educational impacts of deforestation, and thus, from the viewpoint of evidence-based policymaking, further research is needed to better understand the consequences of deforestation.

Notes

1. Another example of such global efforts is Reducing Emissions from Deforestation and forest Degradation (REDD+) by the UN-REDD Programme established in 2007. REDD + aims to reduce carbon emissions by financing developing countries to promote greenhouse gas removal through reducing deforestation and forest degradation.
2. Between 2010 and 2020, Indonesia had the third-largest annual net forest loss of any country, behind Brazil and the Democratic Republic of the Congo (FAO, 2020).
3. For the case of the Amazon forest, there are several studies in the medical and biological literature that suggest a positive relationship between deforestation and malaria incidence (Chaves, Conn, López, & Sallum, 2018; Hahn, Gangnon, Barcellos, Asner, & Patz, 2014; MacDonald & Mordecai, 2019; Olson, Gangnon, Silveira, & Patz, 2010). However, quantitative analyses in these studies, except for MacDonald and Mordecai (2019), are based on coarse empirical strategies: They use data aggregated at the level of a certain administrative unit (in most cases, cross-sectional data) and do not even control for administrative-level fixed effects, which could lead to the deforestation variable being cofounded with various unobserved factors at the administrative level. In addition, Bauhoff and Busch (2020) found no significant associations between deforestation and malaria prevalence by (pooled) cross-section analysis using data from diagnostic test results of rural children in 17 sub-Saharan countries.
4. In addition, the resolution of the satellite imagery is higher for our data: 250 m square for the data they used, compared to 30 m square for the data we used, as explained in Section 3.2.
5. According to the administrative boundaries published by Statistics Indonesia (Badan Pusat Statistik), the average district area size is 3882.3 km^2 and average subdistrict size is 271.8 km^2 in 2014.
6. Some empirical studies have documented the effects of early-life exposure to malaria on educational attainment in adulthood. See, for instance, Barreca (2010), Cutler et al. (2010), Lucas (2010), and Venkataramani (2012). Note also that none of them controlled for sibling (household) fixed effects.

7. For example, experiences with hookworm infection during childhood could reduce wages (Bleakley, 2007); being in good health throughout childhood leads to an increase in adult family incomes (Smith, 2009). In addition, several studies have documented that health improvement can lead to higher educational attainment among children. See, for instance, Bleakley and Lange (2009), Case and Paxson (2010), Case, Fertig, and Paxson (2005), and Ito and Tanaka (2018).
8. In this regard, Koh and Wilcove (2008) used bird and butterfly diversity data to examine how the clearing of forests affect biodiversity and found that 73–83 per cent of the biodiversity loss could be attributed to the change of either primary or secondary forests into oil palm plantations.
9. In Indonesia, eight rounds of the DHS are available (DHS 1987, 1991, 1994, 1997, 2002–2003, 2007, 2012, and 2017). Of these eight, we use the sixth and seventh rounds (i.e. DHS 2007 and 2012). The reason for not using the latest round (DHS 2017) is that it does not contain information on subdistricts (*Kecamatan*) or GPS coordinates, so the respondents' place of residence is only known at the province level. Regarding the earlier rounds of the DHS, as explained later, tree cover data are available only from 2001 onwards, and it is therefore not feasible to incorporate data from the earlier rounds in our analysis. In addition, DHS 2002–2003 has the limitation of not covering the provinces of Nanggroe Aceh Darussalam, Maluku, North Maluku, and Papua due to political instability.
10. After excluding subdistricts in Java and the Lesser Sunda Islands that originally had negligible forest cover and, hence, experienced little forest loss from 2001 onwards, our DHS sample contains 1,540 subdistricts. Note that we use the subdistrict boundaries in 2014 since the administrative boundaries in Indonesia have changed over time.
11. Note that some previous studies, such as Chakrabarti (2021) and Garg (2019), used the forest data from Burgess et al. (2012). We use Hansen et al. (2013) dataset in this study for two main reasons. The first is the data coverage period. While Burgess et al.'s dataset provides information regarding forest loss from 2001 to 2008, Hansen et al.'s dataset covers the period of our focus, that is, from the 2000s to the early 2010s. The second reason is data resolution. While Burgess et al.'s dataset has a resolution of 250 metres, the resolution of Hansen et al.'s dataset is 30 m. Thus, Hansen et al.'s dataset enables us to construct tree cover data with a longer time period and a higher degree of accuracy. Note also that Hansen et al.'s dataset has been used in many studies in both the natural and social science literature (see, for instance, Berazneva & Byker, 2017, 2022; Donaldson & Storeygard, 2016; Yamamoto, 2023; Yamamoto, Shigetomi, Ishimura, & Hattori, 2019).
12. It is also worth noting that the precise definition of deforestation is a critical issue. Tree cover loss data in this study may include logging in existing plantations and agricultural tree crop lands because it is difficult to distinguish them from natural forests in satellite-derived data. Hansen et al.'s dataset mitigates this issue by reporting tree cover loss only in the areas where tree cover was initially observed in 2000, which may exclude the logging and growth cycles in plantations and agricultural land.
13. The DHS 2007 was conducted from July to December, while the DHS 2012 was conducted from January to August.
14. For example, the forest loss during the last 12 months for the respondents who were surveyed in August 2007 is calculated as a weighted average, $Tloss_{j,2006} \times 5/12 + Tloss_{j,2007} \times 7/12$. Similarly, we also adjust the annual forest cover and lagged annual forest loss variables.
15. To retain the large sample size, missing values for these control variables are replaced by the mean values of each variable, and dummy variables for the missing values are also included. An alternative way of dealing with missing values is to estimate using the sample with full information. We confirmed that there is no significant difference between the main results presented below and the results excluding respondents with any missing values.
16. The administrative divisions in Indonesia are as follows: 34 provinces (first level), ~500 regencies and cities (second level), ~7000 subdistricts (third level), and more than 80,000 urban and rural villages (fourth level).
17. The sibling comparison strategy has such an advantage, but it also has the disadvantage of loss of information due to reduced sample size, since children without siblings are not used in the estimation. While we assume that whether or not children have their siblings in the target age group is random, the potential drawbacks of this strategy are discussed in Section 4.2.
18. Supplementary Appendix A.3 provides a more detailed explanation of the sources of these region-level variables.
19. The rapid regency/city splits starting in 1999 were primarily aimed at improving equity and fostering prosperity among local populations (BAPPENAS & UNDP, 2008). On the one hand, considering an undesirable link between deforestation and political factors (Burgess et al., 2012), poorer regions may have faced greater forest extraction. Therefore, our estimates may be biased due to these relationships when not controlling for the number of regencies/cities.
20. The average tree cover loss during the last 12 months in the sample is 5282 thousand square metres and the standard deviation of tree cover loss during the same period is 15,140 thousand square metres. Therefore, $\partial Fever/\partial Tloss_{last12m} \times S.D.(Tloss_{last12m}) = \beta_{last12m} \times S.D.(Tloss_{last12m})/Tloss_{last12m} = 0.0387 \times 15,140/5282 \approx 0.111$. Since the sample average of the incidence of fever is 33 per cent (in the pooled sample), this magnitude amounts to an ~33.6 per cent increase in the incidence of fever.

21. Since the primary symptoms of malaria often include fever, headache, and chills and considering that other potential causes of fever like dengue, chikungunya, and Zika are not strongly associated with tropical deforestation (refer to Section 2.2), the absence of a significant correlation with incidents of cough and diarrhoea may strengthen the case for a potential link between tree cover loss and increased malaria transmission. In this relation, forest fire is one factor contributing to deforestation in Indonesia. While forest fires could increase the incidence of cough, our results show insignificant effects for cough. The reason for this finding may be the time lag in our tree cover loss variable which is measured for the past 12 months, or the weak correlation between our variable and forest fires.
22. The average tree cover loss during the preschool-age period in the DHS sample is 27,134 thousand square metres, and the standard deviation of annual subdistrict-level tree cover loss in the sample (from 2001 to 2012) is 8373 thousand square metres. Therefore, $\partial Educ/\partial Tloss_{pre} \times S.D.(Tloss_t) = \beta_{pre} \times S.D.(Tloss_t)/Tloss_{pre} = -0.169 \times 8373/27,134 \approx -0.052$. Likewise, the average tree cover loss during the school-age period in the sample is 13,960 thousand square metres, and therefore, $\partial Educ/\partial Tloss_{sch} \times S.D.(Tloss_t) = \beta_{sch} \times S.D.(Tloss_t)/Tloss_{sch} = -0.111 \times \frac{8373}{13,960} \approx -0.067$. Likewise, when calculating the magnitude for the IFLS results, we have $\partial Educ/\partial Tloss_{pre} \times S.D.(Tloss_t) = \beta_{pre} \times S.D.(Tloss_t)/Tloss_{pre} = -0.119 \times 10,099/12,113 \approx -0.099$, and $\partial Educ/\partial Tloss_{sch} \times S.D.(Tloss_t) = \beta_{sch} \times S.D.(Tloss_t)/Tloss_{sch} = -0.036 \times 10,099/9043 \approx -0.042$. Thus, we see that the magnitudes of both estimates are in comparable ranges.
23. Note, however, that the sample children in the health and education analyses are different and, hence, caution should be exercised in interpreting their results together.
24. We also examined whether the impacts of deforestation differ by gender. In developing countries, demand for children's education often varies by gender, due to gender differences in roles in the home and labour market (Dhital, Ito, Kaneko, Komatsu, & Yoshida, 2022; Levison, Moe, & Marie Knaul, 2001). The estimation results, however, indicate no heterogeneous effects between genders.
25. The sample size of the test score analysis is 243 less than that of the years of education analysis, but this is not attributed to tree cover loss during the preschool- and school-age periods. Additional analysis not reported here shows that tree cover loss variables have no significant impact on the probability of taking exam.

Acknowledgements

We would like to express our gratitude to Koji Yamazaki and the other participants at the seminar at Kobe University for their helpful comments.

Author contributions

All authors contributed to the study conception and design, data collection, and analysis. Yuki Yamamoto acquired research funding. The first draft of the manuscript was written by Takaaki Kishida and Takahiro Ito. All authors were involved in subsequent draft revisions and approved the final manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by JSPS KAKENHI Grant numbers 19H04340 and 23H03607.

ORCID

Takahiro Ito http://orcid.org/0000-0003-1002-2214
Yuki Yamamoto http://orcid.org/0000-0002-6222-4566

Data availability statement

The data used in this study are from publicly available databases, and the codes are available from the authors upon request.

References

Austin, K. G., Schwantes, A., Gu, Y., & Kasibhatla, P. S. (2019). What causes deforestation in Indonesia? *Environmental Research Letters*, *14*(2), 024007. doi:10.1088/1748-9326/aaf6db

BAPPENAS & UNDP (2008). *Evaluation of the proliferation of administrative region in Indonesia 2001–2007*. Jakarta: BRIDGE (Building and Reinventing Decentralised Governance).

Barreca, A. I. (2010). The long-term economic impact of *in utero* and postnatal exposure to malaria. *Journal of Human Resources*, *45*(4), 865–892. doi:10.1353/jhr.2010.0027

Bauhoff, S., & Busch, J. (2020). Does deforestation increase malaria prevalence? Evidence from satellite data and health surveys. *World Development*, *127*, 104734. doi:10.1016/j.worlddev.2019.104734

Berazneva, J., & Byker, T. S. (2017). Does forest loss increase human disease? Evidence from Nigeria. *The American Economic Review*, *107*(5), 516–521. doi:10.1257/aer.p20171132

Berazneva, J., & Byker, T. S. (2022). Impacts of environmental degradation: Forest loss, malaria, and child outcomes in Nigeria. *Review of Economics and Statistics*, *106*(5), 1254–1267. doi:10.1162/rest_a_01238

Bleakley, H. (2007). Disease and development: Evidence from hookworm eradication in the American south. *The Quarterly Journal of Economics*, *122*(1), 73–117. doi:10.1162/qjec.121.1.73

Bleakley, H. (2010a). Health, human capital, and development. *Annual Review of Economics*, *2*(1), 283–310. doi:10.1146/annurev.economics.102308.124436

Bleakley, H. (2010b). Malaria eradication in the Americas: A retrospective analysis of childhood exposure. *American Economic Journal: Applied Economics*, *2*(2), 1–45. doi:10.1257/app.2.2.1

Bleakley, H., & Lange, F. (2009). Chronic disease burden and the interaction of education, fertility, and growth. *The Review of Economics and Statistics*, *91*(1), 52–65. doi:10.1162/rest.91.1.52

Burgess, R., Hansen, M., Olken, B. A., Potapov, P., & Sieber, S. (2012). The political economy of deforestation in the tropics. *The Quarterly Journal of Economics*, *127*(4), 1707–1754. doi:10.1093/qje/qjs034

Case, A., & Paxson, C. (2010). Causes and consequences of early-life health. *Demography*, *47 Suppl*(Suppl 1), S65–S85. doi:10.1353/dem.2010.0007

Case, A., Fertig, A., & Paxson, C. (2005). The lasting impact of childhood health and circumstance. *Journal of Health Economics*, *24*(2), 365–389. doi:10.1016/j.jhealeco.2004.09.008

Chakrabarti, A. (2021). Deforestation and infant mortality: Evidence from Indonesia. *Economics and Human Biology*, *40*, 100943. doi:10.1016/j.ehb.2020.100943

Chaves, L. S. M., Conn, J. E., López, R. V. M., & Sallum, M. A. M. (2018). Abundance of impacted forest patches less than 5 km^2 is a key driver of the incidence of malaria in Amazonian Brazil. *Scientific Reports*, *8*(1), 7077. doi:10.1038/s41598-018-25344-5

Cohen, J., & Dupas, P. (2010). Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment. *Quarterly Journal of Economics*, *125*(1), 1–45. doi:10.1162/qjec.2010.125.1.1

Corley, R. (2009). How much palm oil do we need? *Environmental Science & Policy*, *12*(2), 134–139. doi:10.1016/j.envsci.2008.10.011

Currie, J., & Almond, D. (2011). Chapter 15 – Human capital development before age five. In *Volume 4 of handbook of labor economics* (pp. 1315–1486). Amsterdam: Elsevier.

Cutler, D., Fung, W., Kremer, M., Singhal, M., & Vogl, T. (2010). Early-life malaria exposure and adult outcomes: Evidence from malaria eradication in India. *American Economic Journal: Applied Economics*, *2*(2), 72–94. doi:10.1257/app.2.2.72

Dhital, R. P., Ito, T., Kaneko, S., Komatsu, S., & Yoshida, Y. (2022). Household access to water and education for girls: The case of villages in hilly and mountainous areas of Nepal. *Oxford Development Studies*, *50*(2), 142–157. doi:10.1080/13600818.2021.1965978

Donaldson, D., & Storeygard, A. (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*, *30*(4), 171–198. doi:10.1257/jep.30.4.171

Dunn, E. C., Soare, T. W., Zhu, Y., Simpkin, A. J., Suderman, M. J., Klengel, T., … Relton, C. L. (2019). Sensitive periods for the effect of childhood adversity on DNA methylation: Results from a prospective, longitudinal study. *Biological Psychiatry*, *85*(10), 838–849. doi:10.1016/j.biopsych.2018.12.023

Edwards, R. B., Naylor, R. L., Higgins, M. M., & Falcon, W. P. (2020). Causes of Indonesia's forest fires. *World Development*, *127*, 104717. doi:10.1016/j.worlddev.2019.104717

FAO (2020). *Global Forest Resources Assessment 2020: Main report*. Rome: Food and Agriculture Organization of the United Nations.

FAO (2022). *FAOSTAT: FAO statistical databases*. Rome: Food and Agriculture Organization of the United Nations.

Garg, T. (2019). Ecosystems and human health: The local benefits of forest cover in Indonesia. *Journal of Environmental Economics and Management*, *98*, 102271. doi:10.1016/j.jeem.2019.102271

Hahn, M., Gangnon, R., Barcellos, C., Asner, G., & Patz, J. (2014). Influence of deforestation, logging, and fire on malaria in the Brazilian Amazon. *PLOS One*, *9*(1), e85725. doi:10.1371/journal.pone.0085725

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., … Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*(6160), 850–853. doi:10.1126/science.1244693

Husnina, Z., Archie, C. A., Clements, A. C. A., & Wangdi, K. (2019). Forest cover and climate as potential drivers for dengue fever in Sumatra and Kalimantan 2006–2016: A spatiotemporal analysis. *Tropical Medicine & International Health*, *24*(7), 888–898. doi:10.1111/tmi.13248

Ito, T., & Tanaka, S. (2018). Abolishing user fees, fertility choice, and educational attainment. *Journal of Development Economics*, *130*, 33–44. doi:10.1016/j.jdeveco.2017.09.006

Kalbus, A., de Souza Sampaio, V., Boenecke, J., & Reintjes, R. (2021). Exploring the influence of deforestation on dengue fever incidence in the Brazilian Amazonas state. *PLOS One*, *16*(1), e0242685. doi:10.1371/journal.pone.0242685

Koh, L. P., & Wilcove, D. S. (2008). Is oil palm agriculture really destroying tropical biodiversity? *Conservation Letters*, *1*(2), 60–64. doi:10.1111/j.1755-263X.2008.00011.x

Levison, D., Moe, K. S., & Marie Knaul, F. (2001). Youth education and work in Mexico. *World Development*, *29*(1), 167–188. doi:10.1016/S0305-750X(00)00090-5

Lucas, A. M. (2010). Malaria eradication and educational attainment: Evidence from Paraguay and Sri Lanka. *American Economic Journal. Applied Economics*, *2*(2), 46–71. doi:10.1257/app.2.2.46

Lucas, A. M. (2013). The impact of malaria eradication on fertility. *Economic Development and Cultural Change*, *61*(3), 607–631. doi:10.1086/669261

MacDonald, A. J., & Mordecai, E. A. (2019). Amazon deforestation drives malaria transmission, and malaria burden reduces forest clearing. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(44), 22212–22218. doi:10.1073/pnas.1905315116

Okiro, E. A., & Snow, R. W. (2010). The relationship between reported fever and *Plasmodium falciparum* infection in African children. *Malaria Journal*, *9*(1), 99. doi:10.1186/1475-2875-9-99

Olson, S., Gangnon, R., Silveira, G., & Patz, J. (2010). Deforestation and malaria in Mâncio Lima county, Brazil. *Emerging Infectious Diseases*, *16*(7), 1108–1115. doi:10.3201/eid1607.091785

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, *37*(2), 187–204. doi:10.1080/07350015.2016.1227711

Pattanayak, S. K., & Pfaff, A. (2009). Behavior, environment, and health in developing countries: Evaluation and valuation. *Annual Review of Resource Economics*, *1*(1), 183–217. doi:10.1146/annurev.resource.050708.144053

Patz, J. A., Graczyk, T. K., Geller, N., & Vittor, A. Y. (2000). Effects of environmental change on emerging parasitic diseases. *International Journal for Parasitology*, *30*(12–13), 1395–1405. doi:10.1016/s0020-7519(00)00141-7

Pimm, S. L., & Raven, P. R. (2000). Biodiversity: Extinction by numbers. *Nature*, *403*(6772), 843–845. doi:10.1038/35002708

Santos, A. S., & Almeida, A. N. (2018). The impact of deforestation on malaria infections in the Brazilian Amazon. *Ecological Economics*, *154*, 247–256. doi:10.1016/j.ecolecon.2018.08.005

Sitohang, V., Sariwati, E., Fajariyani, S. B., Hwang, D., Kurnia, B., Hapsari, R. K., … Hawley, W. A. (2018). Malaria elimination in Indonesia: Halfway there. *The Lancet. Global Health*, *6*(6), E604–E606. doi:10.1016/S2214-109X(18)30198-0

Smith, J. P. (2009). The impact of childhood health on adult labor market outcomes. *The Review of Economics and Statistics*, *91*(3), 478–489. doi:10.1162/rest.91.3.478

Venkataramani, A. S. (2012). Early life exposure to malaria and cognition in adulthood: Evidence from Mexico. *Journal of Health Economics*, *31*(5), 767–780. doi:10.1016/j.jhealeco.2012.06.003

Wang, Y., Hollingsworth, P. M., Zhai, D., West, C. D., Green, J. M., Chen, H., … Ahrends, A. (2023). High-resolution maps show that rubber causes substantial deforestation. *Nature*, *623*(7986), 340–346. doi:10.1038/s41586-023-06642-z

WHO (2021). *World malaria report 2021*. Geneva: World Health Organization.

Wicke, B., Sikkema, R., Dornburg, V., & Faaij, A. (2011). Exploring land use changes and the role of palm oil production in Indonesia and Malaysia. *Land Use Policy*, *28*(1), 193–206. doi:10.1016/j.landusepol.2010.06.001

Yamamoto, Y. (2023). Living under ecosystem degradation: Evidence from the mangrove–fishery linkage in Indonesia. *Journal of Environmental Economics and Management*, *118*, 102788. doi:10.1016/j.jeem.2023.102788

Yamamoto, Y., Shigetomi, Y., Ishimura, Y., & Hattori, M. (2019). Forest change and agricultural productivity: Evidence from Indonesia. *World Development*, *114*, 196–207. doi:10.1016/j.worlddev.2018.10.001

Yasuoka, J., & Levins, R. (2007). Impact of deforestation and agricultural development on anopheline ecology and malaria epidemiology. *The American Journal of Tropical Medicine and Hygiene*, *76*(3), 450–460.

Supplementary Material

Appendix A: Data and Sample

A.1 IFLS Data for Education Analysis

As explained in Section 3.1, we also use the Indonesian Family Life Survey (IFLS) data for the education analysis. This appendix describes the IFLS data in more detail. Out of the five available waves of the IFLS, we utilize the latest wave conducted in 2014-15 (IFLS-5) to examine the causal relationship between forest loss during the preschool- and school-age periods and their subsequent educational attainment.[1] Since tree cover loss data are available from 2001 onwards, only IFLS-5 provides a sufficiently large sample of school-aged children to be matched with the tree cover data. Note that in the IFLS data, the respondents' residence is also known only up to the subdistrict (*Kecamatan*) level, and therefore, the unit of treatment (i.e., tree cover loss) is the subdistrict. The IFLS dataset contains detailed information on individual demographic characteristics and educational outcomes. In the analysis, we employ the highest grade completed by the children and their cognitive test scores. The cognitive test that we use in this study is the test module (named EK1) for children aged 7 to 14, which comprises 12 shape-matching problems and five numeracy problems. In the analysis, we combine the scores of both the shape-matching and the numeracy problems. (See Appendix B.7 for the results for disaggregated test scores.)

Then, we describe the advantages and disadvantages of the DHS and IFLS in the education analysis. A major difference between the DHS and IFLS is in their study areas.

[1] There are five survey rounds, conducted in 1993 (IFLS-1), 1997 (IFLS-2), 2000 (IFLS-3), 2007-08 (IFLS-4), and 2014-15 (IFLS-5).

Although Kalimantan and Sumatra islands have experienced significant deforestation in the recent decades, the IFLS covers only South Kalimantan out of the five provinces in Kalimantan, and also only about half of the provinces in Sumatra. Therefore, in terms of estimating the impact of deforestation, analysis using the DHS would be preferable. On the other hand, due to the limited availability of forest data (as explained in the response to Comment 3a above), the age of the sample children in the DHS 2012 should be between 7 and 11 years, two to three years shorter than those in the IFLS-5 conducted in 2014-2015. This two- or three-year shortness may be problematic when applying the sibling fixed-effects estimation, as a large proportion of the sample children (those without a sibling in the sample) are dropped. In Appendix B.8, we check potential problems due to the reduced sample size.

A.2 Sample and Summary Statistics in the Main Analyses

The summary statistics for the main empirical variables in the child health analysis are reported in Panel A of Table A1. The sample used in the analysis comprises children aged 7 to 59 months old, and we have 10,031 (DHS 2007) and 10,412 (DHS 2012) children after excluding 1,118 (or 5.19%) children with no information on fever.[2] As the health literature suggests, the immune system and nutritional status of newborns and infants, particularly those aged six months and under, are different from those of children older than six months. This difference comes from the intake of breast milk because breastfeeding plays a crucial role in protecting against infections by providing both specific and nonspecific immune factors (Martorell, 1999; Oddy, 2001; Müller and

[2] Note that the total number of children aged 7 to 59 months is 10,624 (DHS 2007) and 10,937 (DHS 2012), and those with no information on cough and diarrhoea are 1,090 (5.06%) and 1,110 (5.15%), respectively.

Krawinkel, 2005; Walters et al., 2016). In Indonesia, approximately 90% of infants up to 6 months old were breastfed during both DHS survey periods.

In addition, we exclude provinces in Java and the Lesser Sunda Islands from our sample,[3] as they originally had negligible forest cover and, hence, the subdistricts in these areas experienced little forest loss from 2001 onwards (see Figure 1). In fact, while the primary forest cover rates in Sumatra and Kalimantan were 34.3% and 56.8%, respectively, the corresponding figures for Java and Bali were less than 1% in 2000 (Margono et al., 2014). Therefore, we exclude these two island areas from the sample, in line with previous empirical studies that have examined deforestation issues in Indonesia (Burgess et al., 2012; Garg, 2019; Yamamoto et al., 2019; Chakrabarti, 2021).

Then, Panel B of Table A1 reports the summary statistics for the main empirical variables for the child education analysis. As explained in Section 3.1, we use the DHS 2012 and IFLS-5 (2014-2015) data for the education analysis. The DHS sample consists of children born between 2001 and 2005 (children aged 7 to 11), and the IFLS sample consists of children born between 2001 and 2008 (children aged 7 to 14). In addition, since the analysis aims to examine the impact of tree cover loss in early childhood on later educational attainment, we restrict the IFLS sample to children from households that did not move out of their subdistrict between 2000 and 2014, while the DHS data do not contain such migration information. Furthermore, as in the health analysis, we also exclude Java and the Lesser Sunda Islands from the sample.[4] Then, after excluding

---

[3] Specifically, we exclude Banten, Banten, Special Capital Region of Jakarta, West Java, Central Java, East Java, and Special Region of Yogyakarta on Java; and Bali, West Nusa Tenggara, and East Nusa Tenggara on the Lesser Sunda Islands.

[4] Among the 13 IFLS provinces, the sample children are living in four provinces on Sumatra (North Sumatra, West Sumatra, South Sumatra, and Lampung) and two provinces on other island groups (South Kalimantan and South Sulawesi).

children without siblings or without information on their years of education or various controls, we are left with 2,685 children in the DHS sample and 775 children in the DHS sample. More specifically, of the 9,971 children aged 7 to 11 in the DHS sample, 951 have no information on their outcome or control variables. In addition, we exclude 6,335 children who have no siblings aged 7 to 11 in the sample due to our strategy of within-sibling comparison, leaving 1,692 children. In addition, of the 1,937 children aged 7 to 14 from households with no migration history in the IFLS sample, 41 have no information on their outcome or control variables. In addition, we exclude 1,121 children who have no siblings in the sample due to our strategy of within-sibling comparison (622 have no siblings in the roster and 499 have no siblings aged 7 to 14), leaving 775 children.

Note also that in the health analysis, there are 329 (1.5%) and 470 (2.2%) cases in the DHS 2007 and 2012 samples, and in the education analysis, there are 104 (1.2%) cases in the DHS sample and 25 (3.2%) cases in the IFLS sample that have no forest loss in any year of the study period. Therefore, we add one to $Floss_{j.}$ when calculating $\ln Floss_{j.}$. To check for sensitivity to this adjustment, we also use different values, such as 10 and 100, but the main results presented later remain virtually unchanged.

Table A1: Summary Statistics

	Mean	Std. Dev.	Min.	Max.
A) Child health	DHS 2007 (# of obs. = 10,031)			
Incidence of disease				
Fever	0.35	0.48	0	1
Cough	0.39	0.49	0	1
Diarrhoea	0.16	0.36	0	1
Forest loss (log)				
During the last 12 months	13.11	2.89	0	19.18
Between 1 and 2 years ago	13.16	2.89	0	18.83
Between 2 and 3 years ago	13.08	2.83	0	18.50
	DHS 2012 (# of obs. = 10,412)			
Incidence of disease				
Fever	0.32	0.47	0	1
Cough	0.35	0.48	0	1
Diarrhoea	0.16	0.36	0	1
Tree cover loss (log)				
During the last 12 months	13.30	2.95	0	19.54
Between 1 and 2 years ago	12.91	3.27	0	18.78
Between 2 and 3 years ago	13.39	3.00	0	19.14
B) Child education	DHS 2012 (# of obs. = 2,685)			
Years of completed education	2.26	1.59	0	6
Tree cover loss (log, per year) during				
Preschool age (6 years or younger)	13.10	2.50	0	19.05
Early preschool age (under 3 years)	12.29	2.49	0	18.64
Late preschool age (3 to 6 years)	13.01	2.53	0	18.78
School age (7 years or older)	13.43	2.59	5.31	19.45
	IFLS-5 (2014-15) (# of obs. = 775)			
Years of completed education	3.46	2.14	0	9
Cognitive test scores (# of obs. = 574)	0.64	0.20	0	1
Tree cover loss (log, per year) during				
Preschool age (6 years or younger)	11.88	2.93	0	17.11
Early preschool age (under 3 years)	11.34	3.44	0	17.23
Late preschool age (3 to 6 years)	11.88	3.18	0	17.04
School age (7 years or older)	11.98	3.07	0	17.69

Notes: This table reports the mean, standard deviation (Std. Dev.), and minimum (Min.) and maximum (Max.) values in DHS 2007 and 2012. The sample in Panel A consists of children aged 7 to 59 months. The sample in Panel B consists of children aged 7 to 14 years.

A.3 Region-Level Data

Precipitation Data: To account for climatic variation across subdistricts, precipitation data are used because, as discussed in Section 2.2, precipitation may be correlated with malaria development and transmission. The precipitation data are from the Climate Hazards Group Infrared Precipitation with Stations (CHIRPS). We use the global monthly information from CHIRPS version 2.0, which is a gridded rainfall time series with a spatial resolution of 0.05 × 0.05 degrees (approximately 5 kilometres × 5 kilometres). The subdistrict-level average monthly precipitation is merged with our sample at the same level.

Population Data: Population and population density may also be correlated with malaria infection because malaria transmission relies on human hosts. Thus, information on population and area at the district regency/city level are used to control for population density. These variables come from the Indonesia Database for Policy and Economic Research (INDO-DAPOER).

Province Data: Deforestation is a result of human-induced economic activity, and economic activity may influence child health and the risk of infectious diseases. In addition, as Burgess et al. (2012) suggested, deforestation was correlated with the introduction of new regencies/cities. We use the survey years' province-level GDP and the number of regencies/cities from the Statistical Yearbook of Indonesia issued by the Statistics Indonesia (GoI, 2008, 2014).[5]

---

[5] Although many empirical studies use nighttime lights as a proxy for the degree of economic development (see, e.g., Chen and Nordhaus (2011) and Donaldson and Storeygard (2016)), there are no reliable and currently available data covering our study period. The DMSP nighttime lights data provided by the US National Oceanic and Atmospheric Administration (NOAA) cover from 1992 to 2013 but were found to be a poor proxy for GDP in Indonesia (Gibson et al., 2021). New NOAA's VIIRS nighttime lights data that are considered to be spatially more accurate are only available from 2012. For these reasons, we rely on the province-level GDP reported by GoI (2008,

Appendix B: Additional Results and Tables

B.1 Correlation Test

Table B1 reports the correlation coefficients between forest loss and *ex ante* child and household characteristics at the subdistrict level. In line with our identification assumption, the partial correlation is calculated after controlling for subdistrict fixed effects to eliminate any between-subdistrict heterogeneity. If no correlations are found, deforestation can be assumed to occur independently from preexisting subdistrict-level trends. The data are from the two DHS rounds (2007 and 2012), and forest loss is measured in terms of the area of deforestation during the *subsequent* 12 months from the time of each survey. The child health outcomes in Panel A are the subdistrict-level percentages of children aged 7 to 59 months old suffering from fever, cough, and diarrhoea (during the two weeks preceding the survey). Child educational attainment in Panel B includes the average years of schooling completed (for children aged 7 to 16) and the primary education completion rate (for children aged 12 to 16). For household characteristics in Panel C, we use parents' education level and the DHS wealth index, which categorizes the sample households into five quintiles: poorest, poorer, middle, richer, and richest.[6] As the table shows, the coefficients are quite small and statistically insignificant for all variables, providing evidence that the occurrence of forest loss is not associated with potential trends in child health and educational outcomes or household affluence after eliminating the time-invariant subdistrict heterogeneity.

2014). Note however that we confirm that employing the DMSP nighttime lights at the subdistrict level instead of proince-level GDP does not change our main results in Section 5.

[6] The DHS wealth index was constructed using data on household possession of a core set of assets, such as televisions, phones, bicycles, and motorcycles; materials used for housing construction, such as wood, ceramic, brick, and cement; and access to different types of drinking water and sanitation facilities.

Table B1: Test of Correlations between Forest Loss and Preexisting Conditions

	(1)	(2)	(3)
	# of subdistricts	Corr. coef.	P-value
A) Incidence of diseases (children aged 7-59 months old)			
Fever	1,928	0.038	0.459
Cough	1,928	−0.033	0.518
Diarrhoea	1,927	−0.032	0.525
B) Schooling			
Years of education completed (children aged 7-16 years)	1,928	−0.023	0.652
Completed primary school (children aged 12-16 years)	1,924	0.028	0.582
C) Household characteristics			
Mother's education level	1,928	0.019	0.702
Father's education level	1,928	−0.001	0.987
Wealth index			
Poorest (1st quintile)	1,928	0.069	0.172
Poorer (2nd quintile)	1,928	0.014	0.779
Middle (3rd quintile)	1,928	−0.036	0.476
Richer (4th quintile)	1,928	−0.048	0.345
Richest (5th quintile)	1,928	−0.019	0.706

Notes: This table reports the results from the correlation tests. Observations are at the subdistrict level. Column 1 reports the number of subdistricts. Column 2 presents the correlation coefficients between forest loss 12 months after the survey and the *ex ante* child health and household wealth statuses conditional on the subdistrict fixed effects. The incidence of disease measures whether a child had been ill during the two weeks preceding the survey. Column 3 reports the p-values.

B.2 Sample with Different Age Groups

In this section, we run regressions similar to those in Table 1 with respect to different age groups. For the main analysis, the sample consists of children aged 7 to 59 months old because of the status of their immune systems. However, it is important to consider whether the estimation results vary if we use different age ranges to construct the sample. For example, many empirical studies on the health effects on infants and young children have focused on children under five years old without accounting for differences in their immune systems by age. Although the period from birth to six months is considered one of the most important periods for child development, it has also been suggested that the age of one year is another important turning point. Hence, we compare our results to those obtained when using the different age groups by estimating the same specification as in Column 3 of Table 1.

Table B2 reports these results: Column 1 presents the estimates for the sample of children aged 0 to 59 months old (under five years). The coefficient on forest loss over the last 12 months decreases slightly from 0.0387 in Column 3 of Table 1 to 0.0347 but remains statistically significant, and the coefficients on forest loss more than one year ago also remain negative and statistically insignificant. Looking at the results in Column 2 for the sample of children aged 13 to 59 months old, we observe that the coefficients are nearly identical to those in the main results (Table 1). Conversely, in Column 3, we see no statistically significant impact of forest loss over the last 12 months for the infant sample, but there is a possible negative association between forest loss two and three years ago and the incidence of fever. As we theorized regarding the reason for the negative relationship between forest loss more than one year ago and the incidence of fever in Section 5.1, deforested areas may become hostile to mosquitoes in the long term, and this may result in a relationship that is stronger for infants than for children under five years old. Overall, the findings in Table B2 indicate that our main results are not sensitive to the choice of age range, such as whether to include infants, and the incidence of infant fever is not affected by forest loss.

Table B2: Estimation with the Sample of Different Age Groups

Dependent variable:	Fever		
Sample: Children aged:	0-59 months old	13-59 months old	0-6 months old
	(1)	(2)	(3)
Forest loss (log)			
During the last 12 months	0.0347*	0.0381**	0.0472
	(0.0178)	(0.0192)	(0.0429)
Between 1 and 2 years ago	−0.0155	−0.0241	−0.0088
	(0.0166)	(0.0185)	(0.0440)
Between 2 and 3 years ago	−0.0234	−0.0183	−0.0773*
	(0.0159)	(0.0169)	(0.0430)
Subdistrict & survey year-month FE	Yes	Yes	Yes
Birth-cohort FE	Yes	Yes	Yes
Individual & household controls	Yes	Yes	Yes
Region-level controls	Yes	Yes	Yes
Observations	23,207	17,897	2,764
Number of subdistricts	1,540	1,540	1,161
R-squared	0.029	0.024	0.136

Notes: Standard errors are clustered at the subdistrict level and reported in parentheses. Statistical significance at the 5 percent level is denoted as ** and at the 10 percent level as *. All columns include the log of the forest cover in the survey year. For brevity, this table presents only the coefficients of interest from Equation (1).

Table B3: Choice of Lag for the Forest Loss Variables

Dependent Variable:	Fever		
(Sample: Children aged 7-59 months old)	(1)	(2)	(3)
Forest Loss (log)			
During the following 12 months			−0.0060
			(0.0145)
During the last 12 months	0.0396**	0.0400**	0.0447**
	(0.0179)	(0.0183)	(0.0215)
Between 1 and 2 years ago	−0.0180	−0.0222	−0.0246
	(0.0176)	(0.0184)	(0.0189)
Between 2 and 3 years ago	−0.0332	−0.0309	−0.0284
	(0.0216)	(0.0216)	(0.0224)
Between 3 and 4 years ago	0.0163	0.0102	0.0087
	(0.0111)	(0.0117)	(0.0120)
Between 4 and 5 years ago		0.0102	0.0087
		(0.0085)	(0.0096)
Subdistrict & survey year-month FE	Yes	Yes	Yes
Birth-cohort FE	Yes	Yes	Yes
Individual & household controls	Yes	Yes	Yes
Subdistrict & regency/city controls	Yes	Yes	Yes
Observations	20,443	20,443	20,443
Number of subdistricts	1,540	1,540	1,540
R-squared	0.027	0.027	0.027

Notes: Standard errors are clustered at the subdistrict level and reported in parentheses. Statistical significance at the 5 percent level is denoted as **. All columns include the log of the forest cover in the survey year. For brevity, this table presents only the coefficients of interest from Equation (1).

B.3 Choice of Lag for the Forest Loss Variables

Table B3 presents the estimation results that include additional forest loss lags—three years before and four years before—in the main specification from Column 3 of Table 1. The aim of this analysis is to determine whether our main results are sensitive to the choice of lag (i.e., whether the lags were subjectively determined according to the results, such as whether they attain statistical significance).

As Columns 1 and 2 of Table B3 indicate, adding these lag variables does not substantially change the estimates in Table 1, suggesting that our choice of lag variables in the main analysis is not problematic. Concerns about subjectivity in the choice of the lags are also ruled out. In addition, forest loss that occurred in the following year is included in this specification to determine whether future forest loss is associated with the incidence of fever because future forest loss should not affect the current incidence of fever. Column 3 adds forest loss during the following 12 months to the regression reported in Column 2. The coefficients on forest loss during the last 12 months and on forest loss between one and two years ago increase slightly in magnitude. However, the main pattern in the results is unchanged, and the coefficient on forest loss during the following 12 months is smaller and insignificant.

B.4 Controlling for Regional Trends in the Child Health Analysis

As discussed in Section 4.2, the health analysis employs subdistrict fixed-effects estimation, which relies on relatively stringent identification assumptions. Therefore, we further address the possible influence of unobserved heterogeneity in child health trends by controlling for regency/city-specific linear time trends. The estimation results in

Column 1 of Table B4 show that the coefficients increase slightly in magnitude relative to those in Table 1, and the effect of tree cover loss over the last 12 months remains positive and statistically significant. Similarly, Columns 2 and 3 show small and insignificant coefficients on the tree cover loss variables for the incidence of cough and diarrhoea. As implied by the robust evidence obtained here, our main results in Table 1 are unlikely to be driven by unobserved region-specific trends over time.

Table B4: Controlling for Region-Specific Linear Trends

Dependent variable:	Fever	Cough	Diarrhoea
(Sample: Children aged 7-59 months)	(1)	(2)	(3)
Tree cover loss (log)			
During the last 12 months	0.0600**	−0.0024	0.0032
	(0.0243)	(0.0283)	(0.0137)
Between 1 and 2 years ago	−0.0259	−0.0019	−0.0159
	(0.0224)	(0.0207)	(0.0140)
Between 2 and 3 years ago	0.0136	−0.0129	−0.0083
	(0.0196)	(0.0222)	(0.0122)
Subdistrict & survey year-month FE	Yes	Yes	Yes
Birth-cohort FE	Yes	Yes	Yes
Individual- & household-level controls	Yes	Yes	Yes
Region-level controls	Yes	Yes	Yes
Observations	20,443	20,471	20,451
Number of subdistricts	1,540	1,540	1,539
R-squared	0.044	0.038	0.052

Notes: Standard errors are clustered at the subdistrict level and reported in parentheses. Statistical significance at the 5 percent level is denoted as **. All columns include the log of the forest cover in the survey year. For brevity, this table presents only the coefficients of interest from Equation (1). "Individual- & household-level controls" include the child age and gender, parental education, and the household wealth index. "Region-level controls” include precipitation at the subdistrict level, population density at the regency/city level, and the number of regencies/cities and GDP at the province level.

B.5 Selection on Observables and Unobservables

In this appendix, we test the validity of our identification assumption by conducting the sensitivity analysis employing the bias-adjusted coefficients proposed by Oster (2019) based on Altonji et al. (2005). If the identification assumption holds true, the baseline effects estimated only with the variables relevant to the identification assumption should be almost unchanged even after including the full set of controls. Conversely, if the estimates are greatly changed, it implies that the issue of selection on unobservables matters. Based on this idea, Oster (2019) proposed a method for calculating the lower (or upper) bound of the treatment effect, as explained below.

The bias-adjusted effects, denoted by β^*, are approximated by (or expressed by, if we assume Assumption 2 from Oster, 2019):

$$\beta^* = \tilde{\beta} - \delta\left(\overset{o}{\beta} - \tilde{\beta}\right)\frac{R_{\max} - \tilde{R}}{\tilde{R} - \overset{o}{R}}.$$

$\tilde{\beta}$ and $\tilde{R}$ are the coefficient estimate of interest and R-squared from a regression with the full set of observed controls (denoted as the *controlled* specification) as in Column 3 of Table 1 in the health analysis and Column 2 of Table 3 in the education analysis. $\overset{o}{\beta}$ and $\overset{o}{R}$ are their equivalents from the *uncontrolled* specification that includes only the variables relevant to the identification assumption (i.e., the subdistrict and survey year-month fixed effects in the health analysis and the sibling and birth year fixed effects in the education analysis). In addition, δ and $R_{\max}$ are parameters that depend on the influences of unobservables and are determined according to the context of the study. δ is defined as the ratio of the selection on unobservables to selection on observables, that is, $\delta = \sigma_{2,X}\sigma_2^{-2}/\sigma_{1,X}\sigma_1^{-2}$, where $\sigma_{j,X}$ is the covariance of the treatment variable with the observables ($j = 1$) or unobservables ($j = 2$), and σ_j is the standard deviation of the observables ($j = 1$) or unobservables ($j = 2$). For δ, we use values ranging from 0.5 to

1.5, indicating that the correlation between the unobservables and treatment status is at most 1.5 times as strong as that between the observables and treatment status. $R_{\max}$ is the regression R-squared assuming that both the observed and unobserved variables can be controlled for. We use two values for $R_{\max}$, assuming that the unobservables explain the variation in the outcome 1.5 and 2 times as much as the controls do (i.e., $R_{\max} = 2.5\tilde{R} - 1.5\overset{o}{R}$ and $R_{\max} = 3\tilde{R} - 2\overset{o}{R}$). Thus, when $\delta = 1.5$ and $R_{\max} = 3\tilde{R} - 2\overset{o}{R}$ together, the influences of unobservables are assumed to be three times as much as those of the observed controls. Table B5 below shows the estimation results from the controlled and uncontrolled specifications.

Then, Figure B1 reports the calculation results for health (Panel A) and educational outcomes (Panel B). The figure shows that the estimates are considerably stable across all the conditions examined regarding the influence of unobservables, except for the coefficients of tree cover loss during preschool age in Panel B.[7] In particular, the estimated impacts of our interest always have the same sign as the estimates in Tables 1 and 3. Thus, even assuming that the influences of unobservables are much larger than those of the observables, we still find that the impact of tree cover loss during the last 12 months on child health and the impact of tree cover loss in the preschool and school-age periods on educational outcomes are not negligible. Therefore, we conclude that our main results are less likely to be driven by possible unobserved confounders.

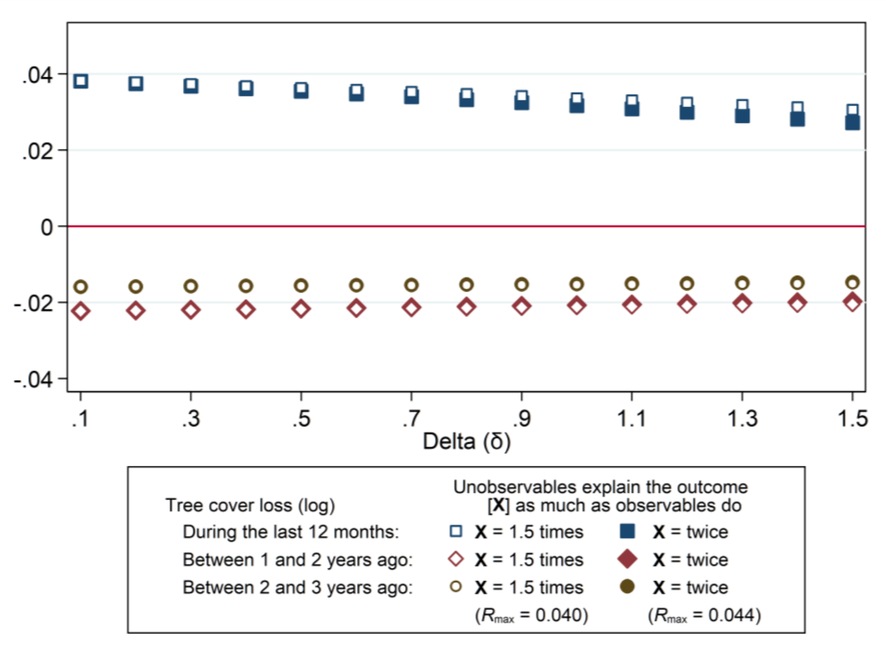
[7] Under the toughest condition, the influences of unobservables are assumed to be three times as much as those of the observable controls.

Table B5: Baseline and Controlled Effects of Forest Loss

	Baseline effect, $\overset{o}{\beta}$	Controlled effect, $\tilde{\beta}$
	(1)	(2)
A) Child health outcomes	Dependent Variable: Fever	
Tree cover loss (log)	(Sample: Children aged 7-59 months old)	
During the last 12 months	0.0390**	0.0387**
	(0.0180)	(0.0183)
Between 1 and 2 years ago	−0.0209	−0.0224
	(0.0175)	(0.0176)
Between 2 and 3 years ago	−0.0159	−0.016
	(0.0163)	(0.0159)
R-square	0.002	0.027
B) Child educational outcomes	Dependent Variable: Years of Education	
Tree cover loss (log)	(Sample: Children aged 7-11 years old)	
During preschool age	−0.182**	−0.213**
	(0.093)	(0.091)
During school age	−0.186**	−0.182**
	(0.084)	(0.082)
R-square	0.804	0.811

Notes: Standard errors are clustered at the subdistrict level and reported in parentheses. Statistical significance at the 1 percent level is denoted as ***, at the 5 percent level as **, and at the 10 percent level as *. The controlled effect is estimated with the full set of controls used in the regression analysis. Panel A presents the results for the child health analysis, and Panel B gives the results for the education analysis.

(A) Fever



(B) Years of Education

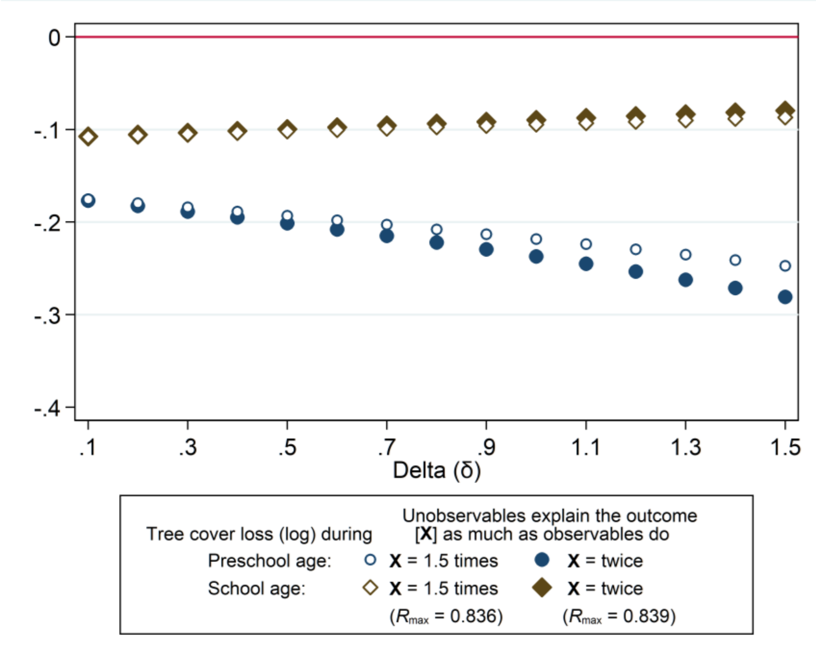


Figure B1: Bias-Adjusted Coefficients of Tree Cover Loss

Notes: The results of Oster's (2019) sensitivity analysis are reported.

B.6 Estimating Educational Impacts using different classifications of exposure period

In this appendix section, we examine the educational consequences of forest loss by further subdividing the age categories. The impact of health shocks can vary greatly with age, and early childhood in particular is considered to be the most vulnerable period. Therefore, we examine heterogeneous effects within the preschool period by dividing the preschool period into two: early preschool (under three years old) and late preschool (three to six years old).

The estimation results are presented in Table B6. The estimation results show that the estimated impact in the under-3 period is smaller for the DHS sample and slightly larger for the IFLS sample than that in the 3 to 6 period. However, the F-statistics for the null hypotheses that they are equal are always small, indicating rejection of the hypotheses (third and fourth rows from the bottom).

B.7 Estimation Results for Disaggregated Cognitive Test Scores

Table B7 reports the estimated impacts of forest loss on shape matching and math test scores separately. Compared with the results in Columns 4 to 6 of Table 6, the main effects on the cognitive test found in Table 6 are through math. Here again, the adverse impact of deforestation is relatively greater during the school age period, suggesting that school absences and poor learning due to illness may be the main causes, rather than impaired cognitive ability during the preschool years.

Table B6: Effects of forest loss using different exposure periods

		DHS		IFLS			
Dependent variable:		Years of education		Years of education		Cognitive test score	
(Sample: Children)		(1)	(2)	(3)	(4)	(5)	(6)
Forest loss (log) during							
Early preschool age (under 3 years) β_1		−0.055	−0.051	−0.075**	−0.068**	−0.013**	−0.007
		(0.046)	(0.046)	(0.027)	(0.027)	(0.006)	(0.006)
Late preschool age (3 to 6 years): β_2		−0.148	−0.158*	−0.051	−0.059	−0.005	−0.005
		(0.092)	(0.092)	(0.043)	(0.040)	(0.008)	(0.008)
School age (7 years or older)		−0.104	−0.108*	−0.036	−0.049	−0.012*	−0.012*
		(0.064)	(0.064)	(0.025)	(0.030)	(0.007)	(0.007)
Sibling & birth-cohort FE		Yes	Yes	Yes	Yes	Yes	Yes
Individual-level controls		Yes	Yes	Yes	Yes	Yes	Yes
Region-level controls		No	No	No	Yes	No	Yes
H_0: $\beta_1 = \beta_2$ (& $\gamma_1 = \gamma_2$);	F-statistic	0.837	0.628	2.107	0.481	0.042	0.407
	(P-value)	0.361	0.429	(0.149)	(0.489)	(0.838)	(0.525)
Observations		2,685	2,685	775	775	532	532
R-squared		0.903	0.903	0.937	0.939	0.739	0.751

Table B7: Effect of Forest Loss on Shape Matching and Math Test Scores

Dependent Variable:	Cognitive Test Score			
	Shape matching		Mathematics	
(Sample: Children aged 7 to 14 years)	(1)	(2)	(3)	(4)
Forest Loss (log) during				
Preschool age (6 years or younger)	−0.002	−0.002	−0.003	−0.003
	(0.013)	(0.014)	(0.015)	(0.013)
School age (7 years or older)	−0.002	−0.005	−0.017	−0.018
	(0.010)	(0.009)	(0.014)	(0.014)
Sibling & Birth-Cohort FE	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes
Subdistrict & Regency/City Controls	No	Yes	No	Yes
Observations	532	532	532	532
R-squared	0.695	0.709	0.684	0.695

B.8 Estimating Educational Impacts with Subdistrict Fixed Effects

Here, we implement subdistrict fixed-effects (not sibling fixed-effects) estimations using the full sample of children aged 7 to 14 as well as the restricted sample used in Table 3 to check for a possible sample selection issue. Columns 1, 3, and 5 of Table B8 report the estimation results using the full sample, and their estimates are expected to contain a bias, if any, due to unobserved heterogeneity across households. The results show that the direction of bias due to unobserved heterogeneity at the household level is mixed: They suggest an upward bias for the educational impact of forest loss during preschool and school age, and a downward bias for cognitive test scores during preschool age.

We then restrict the sample to those with siblings in the target age range (Columns 2, 4, and 6). If the presence of siblings is correlated with the forest loss variables, the estimates should change substantially. The coefficient estimates are relatively stable, suggesting that there is no evidence that restricting the sample to children with siblings biases the estimates. In summary, the main results in Table 3 are not driven by sample restriction, to say the least.

Table B8: Effect of Forest loss on Educational Attainment (with Subdistrict Fixed Effects)

	DHS		IFLS			
Dependent Variable:	Years of Education		Years of Education		Cognitive Test Score	
	(1)	(2)	(3)	(4)	(5)	(6)
Forest Loss (log) during						
Preschool age	−0.128***	−0.097	−0.088*	−0.081*	−0.011	−0.018*
	(0.043)	(0.087)	(0.047)	(0.044)	(0.018)	(0.011)
School age	−0.025	−0.051	−0.043***	−0.044	−0.000	−0.000
	(0.041)	(0.071)	(0.014)	(0.030)	(0.005)	(0.010)
Subdistrict & birth-cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Region-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Whole children aged 7-11	Children with siblings in the sample (Table 3)	Whole children aged 7-14	Children with siblings in the sample (Table 3)	Whole children aged 7-14	Children with siblings in the sample (Table 3)
Observations	9,020	2,685	1,896	775	1,459	571
R-squared	0.727	0.792	0.845	0.856	0.287	0.410

References

Altonji JG, Elder TE, Taber CR (2005) Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. Journal of Political Economy, 113(1): 151–184.

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

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

Chen X, Nordhaus WD (2011) Using luminosity data as a proxy for economic statistics. Proceedings of the National Academy of Sciences, 108(21): 8589–8594.

Donaldson D, Storeygard A (2016) The view from above: Applications of satellite data in economics. Journal of Economic Perspectives, 30(4): 171–98.

Garg T (2019) Ecosystems and human health: The local benefits of forest cover in Indonesia. Journal of Environmental Economics and Management, 98: 102271.

Gibson J, Olivia S, Boe-Gibson G, Li C (2021) Which night lights data should we use in economics, and where? Journal of Development Economics, 149: 102602.

GoI [Government of Indonesia] (2008) Statistical yearbook of Indonesia 2008. Jakarta: Statistics Indonesia (Badan Pusat Statistik).

GoI [Government of Indonesia] (2014). Statistical yearbook of Indonesia 2014. Jakarta: Statistics Indonesia (Badan Pusat Statistik).

Margono BA, Potapov PV, Turubanova S, Stolle F, Hansen MC (2014) Primary forest cover loss in Indonesia over 2000-2012. Nature Climate Change, 4: 730.

Martorell R (1999) The nature of child malnutrition and its long-term implications. Food and Nutrition Bulletin, 20(3): 288–292.

Müller O, Krawinkel M (2005) Malnutrition and health in developing countries. CMAJ, 173(3): 279–286.

Oddy W (2001) Breastfeeding Protects against illness and infection in infants and children: A review of the evidence. Breastfeeding Review, 9(2): 11–18.

Oster E (2019) Unobservable selection and coefficient stability: Theory and evidence. Journal of Business and Economic Statistics, 37(2): 187–204.

Walters D, Horton S, Siregar AYM, Pitriyan P, Hajeebhoy N, Mathisen R, Phan LTH, Rudert C (2016) The cost of not breastfeeding in Southeast Asia. Health Policy and

Planning, 31(8): 1107–1116.

Yamamoto Y, Shigetomi Y, Ishimura Y, Hattori M (2019) Forest change and agricultural productivity: Evidence from Indonesia. World Development, 114: 196–207.