Center for Social Research and Data Archives, Institute of Social Science, The University of Tokyo

SUSTAINABLE DEVELOPMENT GOALS

CSRDA supports the Sustainable Development Goals

# CSRDA Discussion Paper

## Deforestation, Child Health and Education: New Evidence from Indonesia

| No. | Date | SDGs |
|---|---|---|
| 4 | Jan. 2021 | 3 GOOD HEALTH AND WELL-BEING; 4 QUALITY EDUCATION; 15 LIFE ON LAND |

Name

Takaaki Kishida, Takahiro Ito, Yuki Yamamoto

# Deforestation, Child Health and Education: New Evidence from Indonesia*

Takaaki Kishida† Takahiro Ito‡ Yuki Yamamoto§

January 31, 2021

**Abstract**

This study examines the effect of forest loss on child health and education in Indonesia, a country with one of the highest deforestation rates in the world. In line with the findings in the medical and biological literature, which have documented the linkage between deforestation and the breeding of malaria-carrying mosquitoes, our estimation results show that expanding 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 more likely to fall behind academically in terms of grade level but not cognitive performance. Various robustness checks suggest that the adverse health and educational effects are driven by forest loss but not other possible preexisting trends or confounders.

**Keywords:** Deforestation, Forest Loss, Forest Ecosystem, Child Health, Malaria, Education, Indonesia

**JEL classification:** Q23, Q56, Q57, I15, I21, O13

---

*We are deeply grateful for the insightful suggestions from Koji Yamazaki, Buddhika Abeysinghe, and the participants at various seminars at Kobe University. This study was supported by the Japan Society for the Promotion of Science (JSPS) (grant number: 19H04340). The views expressed in this article are those of the authors and do not necessarily reflect the institution.

†Graduate School of International Cooperation Studies, Kobe University. takaaki.kishida.9@gmail.com

‡Graduate School of International Cooperation Studies, Kobe University. takahiro.ito@lion.kobe-u.ac.jp

§Graduate School of Fisheries and Environmental Sciences, Nagasaki University. y-yamamoto@nagasaki-u.ac.jp

# 1 Introduction

Elucidating the cost of environmental devastation precisely is the first and key step in implementing effective environmental policies towards a sustainable society. History has shown, time and again, that industrialization often occurs at the cost of the environment, and worsening environmental conditions have posed severe risks to human life, including epidemics of deadly diseases. In recent years, the international community has made consistent efforts to address these issues, as various targets named in the Sustainable Development Goals (SDGs) are related to environmental issues. However, developing countries, where industrialization is the nation's top priority and enactment of the relevant legislation is delayed, often face difficulties in preventing environmental destruction and protecting people's health.[1] Therefore, rigorous assessment of the consequences of environmental disruption is imperative for evidence-based policymaking.

In this study, we shed light on a less addressed environment-related issue, the health effects of deforestation, by exploiting the case of Indonesia, which has one of the world's highest deforestation rates and has experienced increasing cases of malaria infection.[2] In the literature, an increasing number of studies have documented adverse health consequences of environmental degradation, and most of them have concentrated on chemical contamination, which often emerges as air or water pollution.[3] On the other hand, deforestation has been recognized as a cause of increased greenhouse gases and biodiversity loss, and its direct health

[1]For example, in many tropical countries, the growing demand for forest products is the greatest driver of deforestation, causing critical environmental issues (Hansen et al., 2013; UNECE and FAO, 2018). Between 2001 and 2015, 27% of global forest loss may have been caused by deforestation through land-use changes for commodity production. In fact, 26% and 24% of forest loss was attributed to forestry and agricultural use, respectively (Curtis et al., 2018).

[2]Between 2010 and 2015, Indonesia saw the second-largest annual net forest loss of any country, just behind Brazil (FAO, 2016). Indonesia also had the seventh-highest increase in malaria cases in the world and the highest in Asia in recent years (UNICEF, 2018; WHO, 2018). Moreover, Indonesia had the eighth largest forest area in the world, and 53% of its national land was forested in 2015. The primary cause of deforestation was the transformation of forests into industrial plantations (FAO, 2016).

[3]Many empirical studies have documented that air pollution affects child health in low- and middle-income countries (see, for example, Jayachandran (2009) and Tanaka (2015)). Graff Zivin and Neidell (2013) and Greenstone and Jack (2015) provided a detailed review of the health effects of deteriorating environmental conditions.

cost has been unknown until relatively recently. The medical and biological literature has found that the clearing of forests has the potential to encourage the breeding of malaria-carrying vectors through changes in local ecosystems and environments (Patz et al., 2000; Yasuoka and Levins, 2007; Pattanayak and Pfaff, 2009). In line with these findings, evidence of the linkage between forest loss and malaria infection has been found in some recent 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 that may have been caused by malaria. Studying the relationships between forest cover and human diseases using data on village-level disease outbreaks in Indonesia, Garg (2019) found that forest cover loss is associated with a higher chance of malaria outbreak.[4]

Taking their findings as the point of departure, this paper provides further evidence on the effect of forest loss on malaria infection in early childhood and subsequent educational attainment in Indonesia using datasets obtained from nationally representative surveys and satellite imagery. Our estimation results show that expanding deforestation significantly increases child fever, one of the typical symptoms of malaria infection. In particular, only forest loss during the preceding 12 months has a significant effect on the increase in child fever. On the other hand, the results from falsification tests show that forest loss has no impact on other child diseases, such as cough and diarrhea. This finding implies that deforestation is unrelated to any unobserved trends in child health other than fever, providing evidence on the linkage between deforestation and increased malaria infection. Moreover, the results from the education analysis indicate that exposure to forest loss in early childhood negatively influences subsequent educational attainment in terms of grade levels. Although we cannot determine the detailed mechanism of the negative educational impact and whether and how long the impact persists due to data limitations, the impact may occur because parents delay the admission of malaria-infected children to primary school or have them repeat the same

[4]In addition, Chakrabarti (2021) investigated the effect of deforestation on infant mortality in Indonesia, and the results suggest that firstborn children have a higher risk of infant mortality than later-born children when exposed to deforestation-induced increases in malaria in utero.

grade.

Furthermore, considering the nonexperimental nature of deforestation, in addition to the falsification tests mentioned above, we conduct various checks to test the validity of our identification assumption. First, we investigate the correlation between forest loss and ex ante household characteristics regarding child health and education. If no correlations are found, deforestation could be considered to occur in a plausibly random manner. Then, we further examine possible confounding channels that make it difficult to identify the causal effect. To investigate the influences from region-level confounders and preexisting trends, we estimate the model based on several specifications with and without region characteristics and region-specific linear time trends, which may potentially create different health trends from region to region. Finally, we estimate the bias-adjusted coefficients proposed by Oster (2019) with a more stringent standard to address the possible bias from unobservables. The results from these checks indicate that the identification assumption is plausible and the adverse health and educational effects are driven by forest loss but not other confounding factors or preexisting trends. In the literature on deforestation and its consequences, a few studies have focused on health costs, but none have documented the impact on subsequent human capital development. Thus, providing new evidence on the health and educational consequences of deforestation by adopting a rigorous empirical approach is one of our main contributions to the literature.

This study also contributes to two distinct strands of the literature. First, this study adds to the literature on human capital development by showing that deforestation can be a potential risk factor threatening child health. Malaria is a leading cause of child death in the developing world (WHO, 2018), and our results suggest that preventive measures are necessary for children living in an area where deforestation has occurred recently. Moreover, our results imply that the cost of deforestation may persist in the future through education, in line with the findings in this field that early-life exposure to health shocks can influence future socioeconomic status by lowering educational attainment, labor productivity, and income (Barreca, 2010; Bleakley, 2010b; Cutler et al., 2010; Lucas, 2010, 2013).[5] Children and infants are often the most vulnerable in a deteriorating environment (United Nations, 2019), and therefore, understanding the impact of environmental disruption on child health is of great importance.

Second, this study is related to the literature on the effect of the large increase in palm oil production on rural economic development in developing countries. The expansion of oil palm plantations, one of the major contributors to deforestation in Indonesia, Malaysia, and other equatorial countries (Koh and Wilcove, 2008; Wicke et al., 2011; Hosonuma et al., 2012; Austin et al., 2019), has been a controversial issue. Focusing on the Indonesian case, while some studies noted that palm oil adoption benefited smallholder farmers by improving living standards, wealth accumulation, and nutrition (Euler et al., 2017; Gatto et al., 2017; Krishna et al., 2017), others found that the expansion also had negative effects in several ways. For example, the expansion of palm plantations in Indonesia causes air and water pollution, decreases agricultural production, and lowers female school enrollment (Marlier et al., 2015; Austin et al., 2019; Gatti et al., 2019; Yamamoto et al., 2019; Yamamoto and Ito, 2020). Our study adds the threat of malaria outbreak to the list of potential negative effects of the growing palm oil industry.

The remainder of this paper proceeds as follows. Section 2 describes the background and prior literature on forest loss, health, medicine, and biology relevant to our study. Section 3 outlines the main data used in our analysis, and Section 4 provides the empirical specification, identification strategy, and summary statistics. Section 5 presents the estimation results for the health and education analysis. The conclusion follows.

[5]A sizable body of literature has documented the profound influence of early-life environment and health conditions on socioeconomic outcomes in later life (Bleakley, 2010a; Almond and Currie, 2011; Currie and Vogl, 2013). For example, experiences with hookworm infection during childhood can reduce wages (Bleakley, 2007), and living in good health conditions throughout childhood leads to an increase in adult family incomes (Smith, 2009). In addition, several studies have reported that health improvement can lead to better educational attainment among children (see, for instance, Case et al. (2005); Bleakley and Lange (2009); Case and Paxson (2010); Ito and Tanaka (2018)).

# 2 Background

## 2.1 Forest Loss in Indonesia

Indonesia had the world's second-greatest annual net forest loss (684 thousand ha), just behind Brazil (984 thousand ha), between 2010 and 2015 (FAO, 2016). The major cause of Indonesian deforestation is the clearing of forests to make way for industrial plantations, mostly palm oil (Austin et al., 2019). Indonesia and Malaysia make up 87% of global palm oil production, and Indonesia's palm oil production has quadrupled since 2000 (Koh et al., 2011; Edwards, 2019). Indonesian deforestation will remain a crucial issue because, by 2050, the demand for palm oil in the international market is expected to reach approximately three to seven times the 2005 level (Corley, 2009; Wicke et al., 2011). In addition, an 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, which increased the number of trees felled and decreased equilibrium prices in wood markets due to the increased competition for bribes among local governments (Burgess et al., 2012).

In the literature, both positive and negative effects of the rapid expansion of oil palm plantations have been documented. This expansion aids in economic development that leads to poverty reduction and consumption gains (Edwards, 2019) and improves living standards, wealth accumulation, and nutrition among smallholder farmers in Indonesia (Euler et al., 2017; Gatto et al., 2017; Krishna et al., 2017). In contrast, negative effects have been confirmed from a broader perspective. Expansion-induced deforestation causes air and water pollution and reduces biodiversity in the region (Foley et al., 2005; Marlier et al., 2015; Huijnen et al., 2016; Gatti et al., 2019). Other negative effects of the expansion include a decrease in agricultural production and female school enrollment (Yamamoto et al., 2019; Yamamoto and Ito, 2020). The degraded environments may take up to fifty years or even more to once again resemble the land that existed before the forest was cleared (Patz et al., 2000).

## 2.2 Potential Environmental Impact on Health and Human Capital

Environmental economists have given high priority to tackling issues related to the effects of environmental degradation, such as air and water pollution, on human health (Graff Zivin and Neidell, 2013; Greenstone and Jack, 2015). In particular, the child health burden of air pollution is the topic of a great deal of literature in developing countries, including Indonesia. Jayachandran (2009) assessed the relationship between the wildfire that took place in 1997 and its effect on child mortality in Indonesia. The results showed that over 15,600 children, infants, and fetuses died due to exposure to air pollution. More recently, two studies investigated the long-term impact of exposure to air pollution from the same 1997 Indonesian forest fire. Rosales-Rueda and Triyana (2019) found that children who were exposed to the 1997 air pollution have lower height and lung capacity. Tan-Soo and Pattanayak (2019) also found that the exposure is negatively associated with child development, leading to lower height in adulthood.[6]

Most literature regarding environmental degradation has focused on the influences of pollution, and few empirical investigations have been devoted to understanding the health burden of forest loss and that of ecological disturbance. Berazneva and Byker (2017) combined Nigerian child health data with forest loss data and examined the impact of forest loss on the incidence of malaria. They found that the previous year's forest loss significantly increased child fever caused by malaria infection.[7] Garg (2019) estimated the relationship between district-level forest cover and disease outbreaks in Indonesia. The findings suggested that if a district experienced a 1 percent decrease in forest cover in a

[6]Other studies have focused on both child and adult health. Rangel and Vogl (2019) studied the impact of in utero exposure to smoke from sugarcane harvest fires on health at birth in Brazil and found a negative effect on birth weight, gestational length, and in utero survival. Sheldon and Sankaran (2017) provided evidence on how Indonesian forest fires affect air quality and Singaporean health outcomes, and the results showed that increased fires in Indonesia led to an increase in polyclinic attendance in Singapore. In addition, Graff Zivin and Neidell (2012) found that air pollution affects agricultural worker productivity, suggesting that environmental protection is an important factor in human capital.

[7]In the paper, the authors implicitly used the incidence of fever as a proxy for malaria infection.

given year, the likelihood of malaria outbreaks among the region increases by 2 percentage points. More specifically, the adverse effect of forest loss on malaria outbreaks was driven by primary forest cover loss.[8] The reason for this specific effect is conjectured to be the difference in ecology because primary forests are richer in biodiversity, and loss of biodiversity can lead to increased malaria infection.

In the medical and biological literature, cross-sectional and panel analyses that used data from only part of a country have tended to find positive correlations between deforestation and malaria infection.[9] In contrast, Bauhoff and Busch (2020) used data on rural children in 17 Sub-Saharan countries in Africa and found no significant associations between deforestation and malaria prevalence. They surmised that this finding arose from the unique African context, where the expansion of agriculturally induced deforestation is slower than that in Asia and Latin America, the previous malaria prevalence is high, and malaria is already endemic regardless of forest change.

Furthermore, many empirical studies have documented the well-known effects of early-life exposure to malaria on not only health conditions in adulthood but also socioeconomic status later in life by lowering literacy, educational attainment, labor productivity, and income (Barreca, 2010; Bleakley, 2010b; Cutler et al., 2010; Lucas, 2010, 2013). While evidence on such important linkages has accumulated, the effect of child malaria on subsequent education caused by deforestation remains unaddressed in the literature.

## 2.3 Malaria Transmission and Ecology

In most cases, malaria transmission occurs when one is bitten by 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

[8]Forests in Indonesia can be categorized into primary (virgin) and secondary (disturbed) forests.

[9]See, for example, Yasuoka and Levins (2007) for a detailed review and statistical analysis and Olson et al. (2010), Hahn et al. (2014), Chaves et al. (2018), Santos and Almeida (2018), and MacDonald and Mordecai (2019) for studies using Brazilian deforestation data.

not treated in time, an infected person may develop kidney failure, mental confusion, coma, and even death. Treatment relies on access to and administration of prescription drugs, and the duration of healing varies depending on the person.[10] Malaria is a major cause of death and malnutrition in children worldwide. In Indonesia, fever is a main manifestation of malaria and other acute infections among children under five (DHS, 2012), and Indonesia exhibited the highest increase in malaria cases in Asia in recent years (UNICEF, 2018; WHO, 2018).

The massive clearing of forests has negative effects on ecological and environmental changes, which are strongly linked with human health. As described in the previous section, deforestation can lead to a rise in malaria transmission by increasing the prevalence of Anopheles mosquitoes.

Patz et al. (2000) and Pattanayak and Pfaff (2009) meticulously investigated the linkage between deforestation and malaria transmission, including anopheline larval development, and they showed that forest loss can affect malaria transmission and development in the following four ways. First, cleared lands after deforestation are normally more illuminated by the sun and tend to form puddles with a more neutral potential of hydrogen, helping develop anopheline larvae. For example, agriculture, which involves standing water, such as paddy cultivation and irrigation, provides new suitable breeding sites for mosquitoes and favors the development of larvae. Second, increased ground temperature due to forest loss fosters an increased rate of mosquito development into adults and an increase in their feeding. Third, biodiversity loss caused by deforestation can reduce or eliminate species that prey on Anopheles mosquitoes, thereby indirectly promoting malaria prevalence. For example, using bird and butterfly diversity data to examine how clearing forests affects biodiversity, Koh and Wilcove (2008) found that 73% to 83% of the biodiversity loss could be attributed to changes in either primary or secondary forests into oil palm plantations. Fourth, clearing forests is labor intensive and is generally associated with human migration, which can promote malaria

[10]See https://www.cdc.gov/parasites/malaria/index.html for details.

transmission. Migrants working in forested areas have lower natural immunity and higher contact with mosquitoes, and they tend to have lower access to medical services (Pattanayak and Pfaff, 2009; Garg, 2019).

# 3 Data Sources

## 3.1 Data for Child Health and Education

To explore the relationship between deforestation and the incidence of child malaria, we employ multiple datasets. Data for child health are from the two rounds of the Indonesian Demographic and Health Survey (DHS) conducted in 2007 and 2012 (DHS, 2007, 2012).[11] In these two rounds, respondents' place of residence can be identified only up to the subdistrict (Kecamatan) level.[12] Therefore, forest loss is measured at the subdistrict level, as explained below. Since the Indonesian DHS is a cross-sectional survey, we construct two-period repeated cross-sectional data. In the health analysis, we employ the occurrence of fever as a proxy for the incidence of malaria since the DHS data contain no information on malaria infection and fever is a typical malaria symptom. We also use cough and diarrhea as other child health outcomes; since they are unrelated to malaria infection, employing these symptoms as the dependent variable serves as a falsification test for the linkage between forest loss and child malaria infection. The incidences of fever, cough and diarrhea in children are based on mothers' subjective answers to the questions of whether

[11] In Indonesia, eight rounds of DHS are available (DHS 1987, 1991, 1994, 1997, 2002-03, 2007, 2012, and 2017). Among them, we use the sixth and seventh rounds (i.e., DHS 2007 and 2012). The latest round (DHS 2017) does not contain information on the subdistrict (Kecamatan) or GPS, and hence, respondents' place of residence is known only up to the province level. In addition, DHS 2002-03 does not cover the provinces of Nanggroe Aceh Darussalam, Maluku, North Maluku, and Papua due to political instability. Furthermore, the number of provinces in Indonesia has changed over time, particularly since the late 1990s. Therefore, we employ DHS 2007 and 2012 as the most comprehensive and latest surveys with detailed information on respondents' places of residence.

[12] The administrative divisions in Indonesia are as follows: 34 provinces (first level), approximately 500 regencies and cities (second level), approximately 7,000 subdistricts (third level), and more than 80,000 urban and rural villages (fourth level).

their child suffered from the symptoms during the two weeks preceding the interview; if yes, the value is one, and otherwise, the value is zero.

Data for children's educational attainment are from the Indonesian Family Life Survey (IFLS), the nationally representative survey covering 13 of the 27 provinces where 83% of the Indonesian population is living. Among the five waves of the survey available,[13] we utilize the latest wave conducted in 2014-15 (IFLS-5) to examine the causal relationship between forest loss during the periods of preschool/school ages and subsequent educational levels. As explained below, the forest loss data are available from 2001, and only IFLS-5 can provide a sufficient sample size of school-age children matched with the forest data. Note that in the IFLS data as well, respondents' living place is known only up to the subdistrict (Kecamatan) level, and therefore, the unit of treatment (i.e., forest 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 and comprises 12 shape-matching problems and five numeracy problems.

## 3.2 Forest Data

To construct longitudinal data on forest loss at the subdistrict level, we use the forest data constructed by Hansen et al. (2013). In their updated dataset, the extent of global tree cover in 2000 and forest loss from 2001 to 2019 at a spatial resolution of 30 meters are available.[14] The tree cover extent in a pixel was defined as a percentage from 0% to 100% (in 10% increments) based on the definition that a "tree" is vegetation that is taller than 5 meters in height. Forest loss in a year was defined as the complete removal of "trees" in the pixel and takes unity if it occurred in the year and zero otherwise. Using these data, we construct the

[13]The surveys were conducted in 1993 (IFLS1), 1997 (IFLS-2), 2000 (IFLS-3), 2007-08 (IFLS-4), and 2014-15 (IFLS-5).

[14]See the information on the website at https://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.7.html.

variables of annual forest cover and loss from 2001 to 2014 at the subdistrict level.

Note, however, that it is impossible to calculate the exact forest cover area and deforested area from the forest data provided by Hansen et al. (2013). For example, suppose that the extent of forest cover in a pixel was 20% in 2000 and that the pixel turned to the area with no "trees" (its forest loss variable takes the value of one) in 2003. Then, we cannot determine the exact deforested land area from 2001 to 2003 but know only that the pixel had no "trees" in 2003. Nevertheless, considering the relatively small spatial resolution ($30 \times 30$ square meters) and the human-induced forest loss in Indonesia, including the expansion of oil palm plantations, deforestation probably did not occur gradually over the years but did occur in the short term. Therefore, we assume that deforestation occurred in the reported year.

Taking the 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 the year. Thus, forest loss in the subdistrict $j$ containing $M$ pixels within its border in the year $t$ is defined as $Floss_{jt} = \sum_{m=1}^{M} Loss_{jtm}$, where $Loss_{jtm}$ takes the value of one if the pixel $m$ in the subdistrict $j$ lost "trees" in the year $t$ and zero otherwise. Then, the annual forest cover is calculated as $Fcover_{jt} = \sum_{m=1}^{M} Cover_{j,m}^{2000} - \sum_{s=2001}^{t} Floss_{js}$, where $Cover_{j,m}^{2000}$ is one if the extent of tree cover in pixel $m$ in subdistrict $j$ was 10% or more in 2000 and takes the value of zero otherwise.

# 4 Econometric Framework

## 4.1 Empirical Specification

This study hypothesizes that forest loss has a potential impact on increasing child malaria infections and lowering subsequent educational attainment. For the child health analysis, we estimate the following equation including lagged variables of forest loss to allow for changes in its influence a few years after the occurrence:

$$H_{ijct} = \alpha + \sum_{s=0}^{S} \beta_s \ln Floss_{j,\,t-s} + \gamma \ln Fcover_{jt} + X_{ijt}\delta + \kappa_j + \lambda_c + \mu_\tau + u_{ijct}, \quad (1)$$

in which $i$ denotes the individual; $j$, subdistrict; $c$, child birth cohort; and $t$, (survey) year.

The outcome, $H_{ijct}$, indicates the incidence of disease (fever, cough, or diarrhea) of a child. As mentioned in Section 3.1, cough and diarrhea are for falsification tests. $\ln Floss_{j,\,t-s}$ is the natural logarithm of forest loss during the previous 12 months from year $t-s$ and $S$ is 2. Thus, we focus on the percentage change in forest loss that occurred in the previous three years.[15] $\ln Fcover_{jt}$ is the log of forest cover in year $t$. Note that these forest variables are constructed from the annual forest variables explained in Section 3.2. Since there are differences in the interview month of the DHS between respondents and between survey rounds,[16] the period of 12 months preceding the survey, for example, does not correspond exactly to the last calendar year. Therefore, to adjust for these differences, the forest variables are calculated as the weighted average of the annual forest variables.[17] $X_{ijt}$ is a vector of child, parent, household, and region characteristics. The list of individual- and household-level controls includes children's age and gender, the partner's education, and the household wealth index. These variables are potentially important determinants of child health outcomes.[18] For region characteristics, we use nighttime lights and precipitation at the subdistrict level and population density at the regency/city level (see Section 4.3 for a discussion on the selection of these region-level controls). In addition, we include the subdistrict fixed effects, $\kappa_j$, to control for unobserved time-invariant characteristics of the subdistricts, such as land area,

[15]In the analysis, we will check the sensitivity to the choice of lag variables of forest loss.

[16]The DHS 2007 was conducted from July to December, while the DHS 2012 was conducted from January to August. Table A.1 in Appendix A shows the summary statistics for the survey months of each round of the DHS.

[17]For example, forest loss during the last 12 months for respondents who were surveyed in August 2007 is the sum of $Floss_{j,\,2006} \times 5/12$ and $Floss_{j,\,2007} \times 7/12$. Similarly, we also adjust the annual forest cover and the lagged annual forest loss variables.

[18]To retain the sample size, missing values of these control variables are replaced by the mean values of each variable, and dummy variables for missing values are further controlled.

elevation, and steepness. $\lambda_c$ and $\mu_\tau$ capture fixed effects attributed to the birth cohort and the year-month of the survey in order to eliminate time-specific common health shocks across all subdistricts at the time of birth and survey. $u_{ijct}$ is an unobserved component.

In analyzing the educational attainment of children, letting $E_{ihcj}$ be an educational outcome of child $i$, household $h$, birth cohort $c$, and subdistrict $j$, we estimate the following:

$$E_{ihcj} = \alpha + \sum_{s_{cj}=1}^{2} \left( \beta_{s_{cj}} \ln Floss_{s_{cj}} + \gamma_{s_{cj}} \ln Fcover_{s_{cj}} \right) + X_i\delta + \lambda_h + \mu_c + u_{ihcj}. \quad (2)$$

$\ln Floss_{s_{cj}}$ and $\ln Fcover_{s_{cj}}$ are (the log of) forest loss and forest cover, respectively, during preschool ages $s_{cj} = 1$ and school ages $s_{cj} = 2$, which vary depending on the birth cohort ($c$) as well as subdistrict ($j$). Therefore, we can estimate the impact of forest loss based on within-sibling comparisons by controlling for sibling (household) fixed effects, $\lambda_h$. Note also that we use the *total* deforestation area for $\ln Floss_{s_{cj}}$ and the average annual forest cover area for $\ln Fcover_{s_{cj}}$ in each period. $X_i$ is the individual characteristics such as sex and birth order, $\mu_c$ represents the birth cohort fixed effects, and $u_{ihcj}$ captures unobserved components.

## 4.2 Summary Statistics

The summary statistics of the main empirical variables for the child health analysis are reported in Panel A of Table 1.[19] The sample used in the analysis comprises children aged 7 to 59 months old, with sample sizes of 10,624 (DHS 2007) and 10,937 (DHS 2012). 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 above that age. This difference comes from the intake of breast milk because breastfeeding plays a crucial role in protecting against infections in favor of specific and nonspecific immune

[19]See Table A.2 in Appendix A for the summary statistics of other control variables.

factors (Martorell, 1999; Oddy, 2001; Müller and Krawinkel, 2005; Walters et al., 2016). In Indonesia, approximately 90% of infants up to 6 months old were breastfed in both DHS survey periods (DHS, 2007, 2012).[20] In addition, we exclude from our sample Java and Lesser Sunda islands, which originally had negligible forest cover, and hence, the subdistricts in these areas experienced little forest loss from 2001 onward (see Figure 1 for the maps of forest loss in 2007 and 2012). 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, as with prior empirical studies that examined deforestation issues in Indonesia (Burgess et al., 2012; Garg, 2019; Yamamoto et al., 2019; Chakrabarti, 2021).

Panel B of Table 1 reports the summary statistics of the main empirical variables for the child education analysis.[21] The sample consists of children born from 2001 to 2008 (aged 6 to 14). Because the education analysis aims to investigate the long-term impact of forest loss on educational attainment, we restrict the sample to children in households that did not move outside the subdistrict between 2000 and 2014. Furthermore, as in the health analysis, we also exclude Java and Lesser Sunda islands from the sample.[22] Then, after further excluding 41 children with no information on years of education and other controls, the sample used in the analysis comprises 1,922 children.

There are also 329 and 470 cases (1.5 and 2.2% of the DHS 2007 and 2012 samples, respectively) and 44 cases (3.2% of the IFLS sample) with no forest loss in any year of the study period, and therefore, we added unity when calculating $\ln Floss_{jt}$. To check the sensitivity of this adjustment, we also use different values, such as 10 and 100, but the main results presented later remain virtually unchanged.

[20] We also estimate Equation (1) including newborns and infants aged under seven months to examine how the results change among the different age groups (Section C.1 in Appendix C). As Table C.1 suggests, the inclusion of children under seven months old does not substantially change our main results.

[21] For the summary statistics of other control variables, see Table A.2 in Appendix A.

[22] 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).

[Table 1 About Here]

[Figure 1 About Here]

### 4.3 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 the occurrence is unrelated to preexisting trends in any health- and education-related outcome at the individual and subdistrict levels. In the education analysis based on sibling comparisons, this condition seems plausible because the timing and degree of exposure to forest loss can be exogeneous between siblings. Thus, we assume that after controlling for sibling and birth cohort fixed effects ($\lambda_h$ and $\mu_c$ in Equation 2), forest loss is orthogonal to any observed and unobserved characteristics.

On the other hand, in the health analysis, assuming the above condition may be overly assertive because the identification relies on subdistrict comparisons. For example, during the study period, deforestation occurred mainly in remote areas. Thus, if deforestation areas have a relatively high rate of infectious diseases and low income level, subdistricts that experienced large deforestation in recent years may have a downward trend in health. Such heterogeneous situations in health trends among subdistricts are partly captured by the subdistrict and survey year-month fixed effects ($\kappa_j$ and $\mu_\tau$ in Equation 1), but it is natural to expect that heterogeneous trends remain to be eliminated, causing the health impact of forest loss to be biased.

Thus, given the nonexperimental nature of deforestation, we need to explore its causal impact more carefully. Here, to check the validity of our identification assumption, we conduct the correlation test (Table 2). The reported figures in the table are the correlation coefficients between forest loss and ex ante child and household characteristics at the subdistrict level after controlling for subdistrict fixed effects to eliminate the between-subdistrict heterogeneity based on the identification assumption. If no correlations are found, deforestation can be considered to occur independently from subdistrict-level preexisting trends. The data are from the two DHS rounds (2007 and 2012), and hence, the forest loss is measured as the area of deforestation during the *following* 12 months from the time of each survey. Child health outcomes in Panel A are the subdistrict-level percentage of children aged 7 to 59 months old suffering from fever, cough, and diarrhea (during the two weeks preceding the survey). Child educational attainment in Panel B is the average years of education (children aged 7 to 16) and primary education completion rate (children aged 12 to 16). For household living standards in Panel C, we employ the DHS wealth index, which categorizes the sample households into five quintiles as poorest, poorer, middle, richer, and richest.[23] As the table shows, the coefficients are considerably small and statistically insignificant for all variables, providing no evidence that the occurrence of forest loss is associated with potential trends in child health and educational outcomes after eliminating time-invariant subdistrict heterogeneity.

[Table 2 About Here]

In the analysis presented in Section 5, we further verify our identification strategy in three ways. First, we examine the influence of region-level unobserved trends (in Sections 5.1 and 5.2). By estimating Equation (1) with and without several region-level (i.e., subdistrict- and regency/city-level) controls, we check the robustness of our estimates. If the coefficient estimates of forest loss change largely according to the inclusion of region-level controls, the treatment status (deforestation) is correlated not just with the controls but probably with other unobservables. As such region-level controls, we use nighttime lights and precipitation at the subdistrict level and population density at the regency/city level.[24] Nighttime lights in an area are often used as a proxy for the degree of economic development in the area (Chen and Nordhaus, 2011; Donaldson and Storeygard, 2016) to capture between-subdistrict heterogeneity due to the affluence and health

[23]The DHS wealth index was constructed using data on a household's possessions 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 types of access to drinking water and sanitation facilities.

[24]Appendix B contains a more detailed description of each original data source for these regional variables.

consciousness of people living there. Precipitation and population density are thought to be the key determinants of infectious diseases, as discussed in Section 2.3. Thus, we indirectly check for potential biases due to heterogeneous situations in child health and educational outcomes among subdistricts by comparing the estimates among several specifications. Note that in the education analysis relying on sibling comparisons with cross-sectional data, we employ the average values of these region-level variables during the period from birth to date. In addition, in the health analysis, we further address the possible influences of unobserved regional heterogeneity in child health trends by controlling for region-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 status (subdistrict-level forest loss), we control for regency/city-specific linear time trends, which are the interaction terms between survey years and regency/city dummies.[25]

Second, we investigate whether deforestation has effects on other diseases than fever by conducting falsification tests in Section 5.1. If the effect on child fever is truly caused by malaria infection due to deforestation, other symptoms unrelated to malaria, such as cough and diarrhea, should not have been influenced. Thus, by applying falsification tests, we verify our hypothesis that forest loss has a significant impact on increasing malaria infection in children.

Finally, we attempt to address the issue of selection on unobservables by applying the method proposed by Oster (2019) in Section 5.3. Even if the coefficients are relatively stable across several specifications, the influence of the remaining selection on unobservables is still somewhat unclear. Calculating the approximation of 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, we examine possible biases due to unobserved confounders.

[25] As described in Section 3.1, a regency/city is an administrative division one level higher than a subdistrict.

# 5 Empirical Results

## 5.1 Effect on Child Health

We start by exploring the health effects of deforestation. Table 3 shows the estimation results for the effect of forest loss on child fever. All specifications include the survey year's forest cover and subdistricts and year-month fixed effects. In addition, reported standard errors are clustered at the subdistrict level, allowing existing correlations within the same subdistricts. Column (1) presents the estimated impacts of forest loss during three periods with no additional individual- and region-level controls. The coefficient of forest loss in the last 12 months is 0.0415 and statistically significant at the 5 percent level, while the coefficients of forest loss between one and two years ago and between two and three years ago are negative and not significant. Notably, these results are not sensitive to the choice of lag variables of forest loss. Even though we use additional lags of forest loss, only forest loss in the last 12 months has a significant positive impact on child fever (see Table C.2 in Appendix C).

We then test the robustness of the coefficients by adding various controls: We control for individual and household characteristics (column 2) and subdistrict and regency/city level variables such as precipitation, nighttime lights, and population density (column 3). The results show that the estimated coefficients are considerably stable even when controlling for a rich set of controls, and only forest loss during the last 12 months has a positive and significant impact on child fever. The preferred estimate in column (3) suggests that a 1 standard deviation increase in forest loss significantly increases the incidence of fever by 11.3 percentage points.[26]

[Table 3 About Here]

[26] The standard deviation of forest loss during the last 12 months is 2.909. Therefore, $\beta_{\text{last12mos}} \times S.D(\log Floss_{\text{last12mos}}) = 0.0389 \times 2.909 \approx 0.113$. Since the sample average of the incidence of fever is 33% (in the pooled sample), the magnitude amounts to an approximately 34.2 percent increase in the incidence of fever.

Then, we perform falsification tests to investigate whether the health effect of forest loss is due to malaria infection. If the increased fever incidence found in Table 3 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 4 reports the estimated impacts of forest loss on cough and diarrhea using the same sample and same set of control variables as in Table 3. Columns (1) to (3) are the results for cough, and columns (4) to (6) are for diarrhea. As shown in the table, all the estimated coefficients of forest loss in the last 12 months are much smaller in magnitude than those in Table 3, and none of them are significantly different from zero. Thus, the results indicate that deforestation is unrelated to any preexisting trends in child disease other than fever, providing evidence on the linkage between forest loss and increased malaria transmission.

In addition, the findings that forest loss significantly affects the incidence of malaria infection in the short term are consistent with the findings in previous empirical studies (Berazneva and Byker, 2017; Garg, 2019). A plausible explanation for this short-term impact is that deforested areas may become an environment where mosquitoes cannot live 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 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 several measures to protect children from malaria, and as a result, the incidence of malaria may have decreased. Although the pathways could not 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.

[Table 4 About Here]

As discussed in Section 4.3, in the health analysis, we further address the possible influences of unobserved regional heterogeneity in child health trends by controlling for regency/city-specific linear time trends. The estimation results in Table 5 show that the coefficients slightly increase in magnitude compared with those in column (3) of Table 3, but the effect of forest loss for the last 12 months remains positive and statistically significant. Similarly, columns (2) and (3) present small and insignificant coefficients of the forest loss variables on the incidence of cough and diarrhea. As implied by the robust evidence after controlling for regency/city-specific linear trends, our main results in Table 3 are unlikely to be driven by unobserved region-specific trends over time.

[Table 5 About Here]

## 5.2 Effect on Child Education

Table 6 shows the estimation results for the effect of forest loss on educational outcomes. All specifications include the (log of) average forest cover during preschool and school ages with sibling and birth year fixed effects, and reported standard errors are clustered at the subdistrict level. Columns (1) to (3) present the impacts of forest loss on the number of years of completed education with different specifications: without additional individual- and region-level controls (column 1), with full controls (column 2), and with an interaction term of forest loss between two periods (column 3).[27]

Without the interaction term of forest loss, the coefficient estimate of forest loss during preschool ages is approximately $-0.09$ and statistically significant at the 1 percent level, while that of forest loss during school ages is small and statistically insignificant. Then, including the interaction term (column 3), the estimates of these forest variables increase in magnitude, and in particular, the impact of forest loss during preschool ages is large and still significant at the 1 percent level.[28] The coefficient estimate indicates that a 1 standard deviation increase in forest loss during the preschool period reduces the level of education by

[27]Note that when calculating the interaction term, both forest loss variables are standardized to have zero means by subtracting their average values for ease of interpreting the coefficients of the non-interacted forest loss variables. Thus, Interaction Term = $\left(\ln Floss_{\text{pre}} - \overline{\ln Floss_{\text{pre}}}\right) \times \left(\ln Floss_{\text{sch}} - \overline{\ln Floss_{\text{sch}}}\right)$.

[28]The reason for the increased magnitude of the effects of forest loss is that subdistricts with larger deforestation areas (higher than average) in a period are more likely to have smaller deforestation areas (less than average) in another period. This so-called mean reversion may attenuate the effects of forest loss in each period, and the inclusion of the interaction term eliminates its influence.

0.395 years on average.[29] The impact of the interaction term is not statistically significant but negative, indicating that the larger the deforested area is during both preschool and school ages, the lower the number of years of education. These results indicate that exposure to forest loss during preschool ages (under seven years old) has a persistent effect on subsequent education. Due to data limitations, we cannot examine the mechanism precisely, but children who had malaria in early childhood may enroll in elementary school late or repeat the same grade.

Turning to the results for the cognitive ability of children (columns 4 to 9), we see no statistically significant impact on the score of the shape-matching test, but the result for the math score in column (9) (with the interaction term) indicates the possibility of a negative relationship between deforestation during school ages and the score. The interaction term also has a negative coefficient. Therefore, malaria infection may interrupt education, although the magnitude of the impact is not large: a 1 standard deviation increase in forest loss during school ages, for example, reduces correct answers on the math test by 14.0 percentage points.[30]

[Table 6 About Here]

### 5.3 Selection on Observables and Unobservables

Finally, we examine the validity of our identification assumption by estimating 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 controls composing the identification assumption are supposed to be almost unchanged even after including the full set of controls. Conversely, if the estimates change to a large extent, this implies that the issue of selection on unobservables matters. Thus, in this section, we

[29] The standard deviation of forest loss during preschool ages is 3.064. Therefore, $\beta_{\text{pre}} \times S.D(\log Floss_{\text{pre}}) = -0.129 \times 3.064 \approx -0.395$.

[30] The standard deviation of forest loss during school ages is 3.192. Therefore, $\beta_{\text{sch}} \times S.D(\log Floss_{\text{sch}}) = -0.044 \times 3.192 \approx -0.140$.

conduct a robustness check by calculating the lower (or upper) bound of the treatment effects following Oster (2019) but apply a more rigorous standard.[31]

Table 7 reports the calculation results for child health (Panel A) and education (Panel B). The coefficient estimates reported as the controlled effect (in column 2) are those estimated with full controls in the health and education analyses, i.e., the estimates reported in column (3) of Table 3 and those in column (3) of Table 6, respectively. Regarding the baseline effect in column (1), we control only for the subdistrict and survey year-month fixed effects in the health analysis and for sibling and birth cohort fixed effects in the education analysis, according to the identification assumptions (they are not reported in Tables 3 and 6).[32] Then, based on these estimates, we calculate the interval of the treatment effects ("Identification Set" in column 3) with Oster's bias-adjusted coefficients.

As the results for the main forest variables show, the bias-adjusted effect of forest loss during the last 12 months on fever (Panel A) and that of forest loss during preschool ages on education (Panel B) are smaller in magnitude than the corresponding controlled effects. However, the identified sets for these forest loss variables exclude zero and are within the 95% confidence interval of the controlled effects. Thus, even assuming that the influence of unobservables is four times that of observed controls, based on a tougher standard in calculating the bias-adjusted treatment effects than those employed in Oster (2019), we still find sizable effects of forest loss on child health and education.

Overall, the various robustness checks presented in this section and Appendix C indicate that the main results are very robust and not driven by possible preexisting trends. Thus, children exposed to larger-scale deforestation in early childhood are more likely to have malaria infection and fall behind academically in terms of grade level (but not in cognitive performance).

[Table 7 About Here]

[31] Specifically, we assume that the unobserved controls explain four times as much of the outcome as the observed controls; that is, we choose $R_{\max} = 3\tilde{R} - 2\mathring{R}$ and $\delta = 2$. See Appendix D for the detailed method.

[32] In both estimations, the forest cover variable(s) are also controlled together with the forest loss variables.

# 6 Conclusion

In this study, we examined the effect of forest loss on child malaria infection and subsequent educational attainment using datasets obtained from nationally representative surveys and satellite imagery. The empirical findings imply that forest loss in the previous twelve months significantly increased child malaria infection and that exposure to deforestation during preschool ages subsequently led to reductions in the number of years of completed education.

These results have important policy implications. First, malaria outbreaks can be triggered by recent forest loss. Therefore, countermeasures need to be taken promptly in areas with ongoing deforestation or under potential deforestation risk. One effective measure would be the free distribution of mosquito nets to households (Cohen and Dupas, 2010). Second, in addition to the health burden, adverse educational consequences need to be included in the cost of deforestation. Moreover, the cost may increase in the future because of the accumulation of the negative impact on the attainment of higher levels of education. Understating the cost of deforestation may result in delayed or poor implementation of countermeasures, and therefore, our findings emphasize the importance of adopting effective measures. Furthermore, considering the increasing global demand for palm oil products, forest conservation continues to be challenging in Indonesia and other equatorial countries with expansions in oil palm plantations (Corley, 2009; Wicke et al., 2011). Viewed in this light, examining the cost of deforestation has become increasingly important.

However, the task of evaluating the impacts of deforestation still leaves several issues unaddressed. For example, the child health outcomes used in the health analysis are subjective data based on mothers' reports on their children's diseases but not actual diagnostic data, as the DHS data used in the analysis contain no information on the results of the medical examination for malaria parasites. In addition, the education analysis in this study employed the sample of school-age children due to the limited forest data available, and therefore, the analysis is silent on whether and how long the adverse impact persists in secondary and tertiary education or in the labor market. The literature on the health and educational impacts of deforestation is surprisingly sparse, and from the viewpoint of evidence-based policymaking, further research addressing these issues is needed to better understand the consequences of deforestation.

# References


Almond, Douglas and Janet Currie. 2011. Human capital development before age five. In *Handbook of Labor Economics*, vol. 4B, chap. 15. Elsevier, 1 ed., 1315–1486.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy* 113 (1):151–184.

Austin, Kemen G, Amanda Schwantes, Yaofeng Gu, and Prasad S Kasibhatla. 2019. What causes deforestation in Indonesia? *Environmental Research Letters* 14 (2):024007.

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

Bauhoff, Sebastian and Jonah Busch. 2020. Does deforestation increase malaria prevalence? Evidence from satellite data and health surveys. *World Development* 127:104734.

Berazneva, Julia and Tanya S. Byker. 2017. Does forest loss increase human disease? Evidence from Nigeria. *American Economic Review* 107 (5):516–21.

Bleakley, Hoyt. 2007. Disease and development: Evidence from hookworm eradication in the American South. *The Quarterly Journal of Economics* 122 (1):73–117.

———. 2010a. Health, human capital, and development. *Annual Review of Economics* 2:283–310.

———. 2010b. Malaria eradication in the Americas: A retrospective analysis of childhood exposure. *American Economic Journal: Applied Economics* 2 (2):1–45.

Bleakley, Hoyt and Fabian Lange. 2009. Chronic disease burden and the interaction of education, fertility, and growth. *The Review of Economics and Statistics* 91 (1):52–65.

Burgess, Robin, Matthew Hansen, Benjamin A. Olken, Peter Potapov, and Stefanie Sieber. 2012. The political economy of deforestation in the tropics. *The Quarterly Journal of Economics* 127 (4):1707–1754.

Case, Anne, Angela Fertig, and Christina Paxson. 2005. The lasting impact of childhood health and circumstance. *Journal of Health Economics* 24 (2):365–389.

Case, Anne and Christina Paxson. 2010. Causes and consequences of early-life health. *Demography* 47 (1):S65–S85.

Chakrabarti, Averi. 2021. Deforestation and infant mortality: Evidence from Indonesia. *Economics & Human Biology* 40:100943.

Chaves, Leonardo, Jan Conn, Rossana Veronica Lopez, and Maria Sallum. 2018. Abundance of impacted forest patches less than 5 km2 is a key driver of the incidence of malaria in Amazonian Brazil. *Scientific Reports* 8.

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

Cohen, Jessica and Pascaline Dupas. 2010. Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment. *The Quarterly Journal of Economics* 125 (1):1–45.

Corley, R.H.V. 2009. How much palm oil do we need? *Environmental Science & Policy* 12 (2):134 – 139.

Currie, Janet and Tom Vogl. 2013. Early-life health and adult circumstance in developing countries. *Annual Review of Economics* 5 (1):1–36.

Curtis, Philip G., Christy M. Slay, Nancy L. Harris, Alexandra Tyukavina, and Matthew C. Hansen. 2018. Classifying drivers of global forest loss. *Science* 361 (6407):1108–1111.

Cutler, David, Winnie Fung, Michael Kremer, Monica Singhal, and Tom Vogl. 2010. Early-life malaria exposure and adult outcomes: Evidence from malaria eradication in India. *American Economic Journal: Applied Economics* 2 (2):72–94.

DHS. 2007. Indonesia demographic and health survey 2007. https://dhsprogram.com/publications/publication-fr218-dhs-final-reports.cfm.

———. 2012. Indonesia demographic and health survey 2012. https://dhsprogram.com/publications/publication-fr275-dhs-final-reports.cfm.

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

Edwards, Ryan B. 2019. Export agriculture and rural poverty: Evidence from Indonesian palm oil. Working Paper.

Euler, Michael, Vijesh Krishna, Stefan Schwarze, Hermanto Siregar, and Matin Qaim. 2017. Oil oalm adoption, household welfare, and nutrition among smallholder farmers in Indonesia. *World Development* 93:219 – 235.

FAO. 2016. The global forest resources assessment 2015: How are the world's forests changing? Rome: Food and Agriculture Organization of the United Nations. ISBN 978-92-5-109283-5.

Foley, Jonathan A., Ruth DeFries, Gregory P. Asner, Carol Barford, Gordon Bonan, Stephen R. Carpenter, F. Stuart Chapin, Michael T. Coe, Gretchen C. Daily, Holly K. Gibbs, Joseph H. Helkowski, Tracey Holloway, Erica A. Howard, Christopher J. Kucharik, Chad Monfreda, Jonathan A. Patz, I. Colin Prentice, Navin Ramankutty, and Peter K. Snyder. 2005. Global consequences of land use. *Science* 309 (5734):570–574.

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

Gatti, Roberto Cazzolla, Jingjing Liang, Alena Velichevskaya, and Mo Zhou. 2019. Sustainable palm oil may not be so sustainable. *Science of The Total Environment* 652:48 – 51.

Gatto, Marcel, Meike Wollni, Rosyani Asnawi, and Matin Qaim. 2017. Oil palm boom, contract farming, and rural economic development: Village-level evidence from Indonesia. *World Development* 95:127 – 140.

Graff Zivin, Joshua and Matthew Neidell. 2012. The impact of pollution on worker productivity. *American Economic Review* 102 (7):3652–73.

———. 2013. Environment, health, and human capital. *Journal of Economic Literature* 51 (3):689–730.

Greenstone, Michael and B. Kelsey Jack. 2015. Envirodevonomics: A research agenda for an emerging field. *Journal of Economic Literature* 53 (1):5–42.

Hahn, Micah, Ronald Gangnon, Christovam Barcellos, Gregory Asner, and Jonathan Patz. 2014. Influence of deforestation, logging, and fire on malaria in the Brazilian Amazon. *PloS ONE* 9:e85725.

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

Hosonuma, Noriko, Martin Herold, Veronique De Sy, Ruth S De Fries, Maria Brockhaus, Louis Verchot, Arild Angelsen, and Erika Romijn. 2012. An assessment of deforestation and forest degradation drivers in developing countries. *Environmental Research Letters* 7 (4):044009.

Huijnen, Vincent, M. Wooster, Johannes Kaiser, D. Gaveau, Johannes Flemming, Mark Parrington, A. Inness, Daniel Murdiyarso, B. Main, and M. van Weele. 2016. Fire carbon emissions over maritime southeast Asia in 2015 largest since 1997. *Scientific Reports* 6:26886.

Ito, Takahiro and Shinsuke Tanaka. 2018. Abolishing user fees, fertility choice, and educational attainment. *Journal of Development Economics* 130:33 – 44.

Jayachandran, Seema. 2009. Air quality and early-life mortality: Evidence from Indonesia's wildfires. *The Journal of Human Resources* 44 (4):916–954.

Koh, Lian Pin, Jukka Miettinen, Soo Chin Liew, and Jaboury Ghazoul. 2011. Remotely sensed evidence of tropical peatland conversion to oil palm. *Proceedings of the National Academy of Sciences* 108 (12):5127–5132.

Koh, Lian Pin and David S. Wilcove. 2008. Is oil palm agriculture really destroying tropical biodiversity? *Conservation Letters* 1 (2):60–64.

Krishna, Vijesh, Michael Euler, Hermanto Siregar, and Matin Qaim. 2017. Differential livelihood impacts of oil palm expansion in Indonesia. *Agricultural Economics* 48 (5):639–653.

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

———. 2013. The impact of malaria eradication on fertility. *Economic Development and Cultural Change* 61 (3):607–631.

MacDonald, Andrew J and Erin A Mordecai. 2019. Amazon deforestation drives malaria transmission, and malaria burden reduces forest clearing: a retrospective study. *The Lancet Planetary Health* 3:S13. Planetary Health Annual Meeting 2019.

Margono, Belinda Arunarwati, Peter V. Potapov, Svetlana Turubanova, Fred Stolle, and Matthew C. Hansen. 2014. Primary forest cover loss in Indonesia over 2000 - 2012. *Nature Climate Change* 4:730.

Marlier, Miriam E, Ruth S DeFries, Patrick S Kim, Shannon N Koplitz, Daniel J Jacob, Loretta J Mickley, and Samuel S Myers. 2015. Fire emissions and regional air quality impacts from fires in oil palm, timber, and logging concessions in Indonesia. *Environmental Research Letters* 10 (8):085005.

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

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

Oddy, Wendy. 2001. Breastfeeding protects against illness and infection in infants and children: a review of the evidence. *Breastfeeding Review* 9 (2):11–18.

Olson, Sarah, Ronald Gangnon, Guilherme Silveira, and Jonathan Patz. 2010. Deforestation and malaria in Mancio Lima county, Brazil. *Emerging Infectious Diseases* 16:1108–15.

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

Pattanayak, Subhrendu K. and Alexander Pfaff. 2009. Behavior, environment, and health in developing countries: Evaluation and valuation. *Annual Review of Resource Economics* 1 (1):183–217.

Patz, Jonathan A., Thaddeus K. Graczyk, Nina Geller, and Amy Y. Vittor. 2000. Effects of environmental change on emerging parasitic diseases. *International Journal for Parasitology* 30 (12):1395 – 1405.

Rangel, Marcos A. and Tom S. Vogl. 2019. Agricultural fires and health at birth. *The Review of Economics and Statistics* 101 (4):616–630.

Rosales-Rueda, Maria and Margaret Triyana. 2019. The persistent effects of early-life exposure to air pollution: Evidence from the Indonesian forest fires. *Journal of Human Resources* 54 (4):1037–1080.

Santos, Augusto Seabra and Alexandre N. Almeida. 2018. The impact of deforestation on malaria infections in the Brazilian Amazon. *Ecological Economics* 154:247 – 256.

Sheldon, Tamara L. and Chandini Sankaran. 2017. The impact of Indonesian forest fires on Singaporean pollution and health. *American Economic Review* 107 (5):526–29.

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

Tan-Soo, Jie-Sheng and Subhrendu K. Pattanayak. 2019. Seeking natural capital projects: Forest fires, haze, and early-life exposure in Indonesia. *Proceedings of the National Academy of Sciences* 116 (12):5239–5245.

Tanaka, Shinsuke. 2015. Environmental regulations on air pollution in China and their impact on infant mortality. *Journal of Health Economics* 42:90 – 103.

UNECE and FAO. 2018. Forest products annual market review 2017-2018 .

UNICEF. 2018. Levels & trends in child mortality: Report 2018, estimates developed by the United Nations inter-agency group for child mortality estimation. United Nations Inter-

Agency Group for Child Mortality Estimation (UNIGME).

United Nations. 2019. About the sustainable development goals. https://www.un.org/sustainabledevelopment/sustainable-development-goals/. Accessed: August 16, 2020.

Walters, Dylan, Susan Horton, Adiatma Yudistira Manogar Siregar, Pipit Pitriyan, Nemat Hajeebhoy, Roger Mathisen, Linh Thi Hong Phan, and Christiane Rudert. 2016. The cost of not breastfeeding in Southeast Asia. *Health Policy and Planning* 31 (8):1107–1116.

WHO. 2018. World malaria report 2018. Geneva: World Health Organization. Licence: CC BY-NC-SA 3.0 IGO.

Wicke, Birka, Richard Sikkema, Veronika Dornburg, and Andre Faaij. 2011. Exploring land use changes and the role of palm oil production in Indonesia and Malaysia. *Land Use Policy* 28 (1):193 – 206.

Yamamoto, Yuki and Takahiro Ito. 2020. Leaving school for marriage? New evidence on the impact of agricultural price boom from Indonesian oil palm industry. Working Paper.

Yamamoto, Yuki, Yosuke Shigetomi, Yuichi Ishimura, and Mitsuru Hattori. 2019. Forest change and agricultural productivity: Evidence from Indonesia. *World Development* 114:196 – 207.

Yasuoka, Junko and Richard Levins. 2007. Impact of deforestation and agricultural development on anopheline ecology and malaria epidemiology. *The American Journal of Tropical Medicine and Hygiene* 76:450–60.

# Tables and Figures

**Table 1:** Summary Statistics

| | Mean | S.D. | Min | Max |
|---|---|---|---|---|
| Panel A. Child Health | DHS 2007 ($N$ = 10,624) | | | |
| Incidence of Diseases | | | | |
| Fever | 0.35 | 0.48 | 0 | 1 |
| Cough | 0.39 | 0.49 | 0 | 1 |
| Diarrhea | 0.16 | 0.36 | 0 | 1 |
| Forest Loss (log) | | | | |
| During the Last 12 Months | 13.122 | 2.877 | 0 | 19.183 |
| Between 1 and 2 Years Ago | 13.170 | 2.878 | 0 | 18.830 |
| Between 2 and 3 Years Ago | 13.092 | 2.817 | 0 | 18.495 |
| | DHS 2012 ($N$ = 10,937) | | | |
| Incidence of Diseases | | | | |
| Fever | 0.32 | 0.47 | 0 | 1 |
| Cough | 0.35 | 0.48 | 0 | 1 |
| Diarrhea | 0.16 | 0.36 | 0 | 1 |
| Forest Loss (log) | | | | |
| During the Last 12 Months | 13.304 | 2.935 | 0 | 19.538 |
| Between 1 and 2 Years Ago | 12.911 | 3.251 | 0 | 18.780 |
| Between 2 and 3 Years Ago | 13.399 | 2.981 | 0 | 19.137 |
| Panel B. Child Education | IFLS-5 (2014-15) ($N$ = 1,922) | | | |
| Educational Attainment | | | | |
| Years of Completed Education | 3.450 | 2.088 | 0 | 9 |
| Cognitive Test Scores ($N$ = 1,460) | | | | |
| Shape Matching | 0.665 | 0.216 | 0 | 1 |
| Mathematics | 0.549 | 0.257 | 0 | 1 |
| Forest loss (log) | | | | |
| During Preschool Ages | 13.979 | 3.064 | 0 | 19.053 |
| During School Ages | 13.336 | 3.362 | 0 | 19.133 |

*Notes*: This table reports the mean, standard deviation (S.D.), and minimum (Min) and maximum (Max) values in the DHS 2007 and 2012. The sample in Panel A consists of children aged 7 to 59 months old. The sample in Panel B consists of children aged 7 to 16 years old.

**Table 2:** Test of Correlations between Forest Loss and Preexisting Conditions

| | (1) # of Subdistricts | (2) Corr. Coef. | (3) P-value |
|---|---|---|---|
| Panel A. Incidence of Diseases (Children Aged 7-59 Months Old) | | | |
| Fever | 1,928 | 0.038 | 0.459 |
| Cough | 1,928 | -0.033 | 0.518 |
| Diarrhea | 1,927 | -0.032 | 0.525 |
| Panel B. Schooling | | | |
| Years of Education | 1,928 | -0.023 | 0.652 |
| (Children Aged 7-16 Years Old) | | | |
| Primary School Completed | 1,924 | 0.028 | 0.582 |
| (Children Aged 12-16 Years Old) | | | |
| Panel C. Household Living Standard (Wealth Index) | | | |
| Poorest | 1,928 | 0.069 | 0.172 |
| Poorer | 1,928 | 0.014 | 0.779 |
| Middle | 1,928 | -0.036 | 0.476 |
| Richer | 1,928 | -0.048 | 0.345 |
| Richest | 1,928 | -0.019 | 0.706 |

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

**Table 3:** Effect of Forest Loss on Fever Incidence

| Dependent Variable | Fever | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Forest Loss (log) | | | |
| During the Last 12 Months | 0.0415** | 0.0395** | 0.0389** |
| | (0.0180) | (0.0174) | (0.0178) |
| Between 1 and 2 Years Ago | -0.0231 | -0.0234 | -0.0244 |
| | (0.0175) | (0.0171) | (0.0173) |
| Between 2 and 3 Years Ago | -0.0164 | -0.0161 | -0.0151 |
| | (0.0163) | (0.0159) | (0.0159) |
| Subdistrict & Survey Year-Month FE | Yes | Yes | Yes |
| Birth Cohort FE | Yes | Yes | Yes |
| Individual & Household Controls | | Yes | Yes |
| Subdistrict & Regency/City Controls | | | Yes |
| Observations | 20,443 | 20,443 | 20,443 |
| Number of Subdistricts | 1,540 | 1,540 | 1,540 |
| R-squared | 0.002 | 0.027 | 0.027 |

*Notes*: Standard errors are clustered at the subdistrict level in parentheses. Statistical significance is denoted as ** at the 5 percent level. All columns include the log of the survey year's forest cover. For brevity, this table presents only the coefficients of interest from Equation (1). The sample in this analysis consists of children aged 7 to 59 months old. Individual and household controls include the children's age and gender, the partner's education, and the household wealth index; precipitation and nighttime lights are at the subdistrict level; and population density is at the district level as explained in Table A.2.

**Table 4:** Falsification Tests: Effect of Forest Loss on Cough and Diarrhea Incidence

| Dependent Variable | Cough | | | Diarrhea | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Forest Loss (log) | | | | | | |
| During the Last 12 Months | 0.00890 | 0.00770 | 0.00677 | 0.000488 | -0.000567 | 0.00189 |
| | (0.0194) | (0.0193) | (0.0195) | (0.0141) | (0.0143) | (0.0142) |
| Between 1 and 2 Years Ago | -0.0231 | -0.0237 | -0.0226 | -0.00130 | -0.00270 | -0.00567 |
| | (0.0194) | (0.0193) | (0.0191) | (0.0119) | (0.0119) | (0.0117) |
| Between 2 and 3 Years Ago | -0.0215 | -0.0213 | -0.0217 | -0.00668 | -0.00466 | -0.00362 |
| | (0.0179) | (0.0178) | (0.0178) | (0.0131) | (0.0133) | (0.0133) |
| Subdistrict & Survey Year-Month FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Birth Cohort FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual & Household Controls | | Yes | Yes | | Yes | Yes |
| Subdistrict & Regency/City Controls | | | Yes | | | Yes |
| Observations | 20,471 | 20,471 | 20,471 | 20,451 | 20,451 | 20,451 |
| Number of Subdistricts | 1,540 | 1,540 | 1,540 | 1,539 | 1,539 | 1,539 |
| R-squared | 0.010 | 0.019 | 0.019 | 0.022 | 0.032 | 0.032 |

*Notes*: Standard errors are clustered at the subdistrict level in parentheses. All columns include the log of the survey year's forest cover. For brevity, this table presents only the coefficients of interest from Equation (1). The sample in this analysis consists of children aged 7 to 59 months old. Individual and household controls include the children's age and gender, the partner's education, and the household wealth index; precipitation and nighttime lights are at the subdistrict level; and population density is at the district level as explained in Table A.2.

**Table 5:** Effect on Disease Incidence: Controlling for Region-Specific Linear Trends

| Dependent Variable | Fever | Cough | Diarrhea |
|---|---|---|---|
| | (1) | (2) | (3) |
| Forest Loss (log) | | | |
| During the Last 12 Months | 0.0594** | 0.00111 | 0.00264 |
| | (0.0238) | (0.0264) | (0.0144) |
| Between 1 and 2 Years Ago | -0.0264 | -0.00445 | -0.0165 |
| | (0.0221) | (0.0202) | (0.0142) |
| Between 2 and 3 Years Ago | 0.0140 | -0.0114 | -0.00732 |
| | (0.0199) | (0.0224) | (0.0122) |
| 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 | 20,471 | 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 in parentheses. Statistical significance is denoted as ** at the 5 percent level. All columns include the log of the survey year's forest cover. For brevity, this table presents only the coefficients of interest from Equation (1). The sample in this analysis consists of children aged 7 to 59 months old. Individual and household controls include the children's age and gender, the partner's education, and the household wealth index; precipitation and nighttime lights are at the subdistrict level; and population density is at the district level as explained in Table A.2.

**Table 6:** Effect of Forest Loss during Preschool and School Age on Educational Outcomes

| Dependent Variable | Years of Education | | | Cognitive Test Scores | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Shape Matching | | | Math | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Forest Loss (log) | | | | | | | | | |
| During Preschool Age | -0.092*** | -0.090*** | -0.129*** | 0.001 | 0.004 | -0.006 | -0.000 | 0.001 | -0.024 |
| | (0.022) | (0.022) | (0.037) | (0.010) | (0.010) | (0.015) | (0.011) | (0.011) | (0.017) |
| During School Age | -0.008 | -0.013 | -0.057 | -0.000 | 0.001 | -0.009 | -0.014 | -0.016 | -0.044** |
| | (0.018) | (0.019) | (0.044) | (0.010) | (0.009) | (0.015) | (0.014) | (0.014) | (0.020) |
| Interaction Term: [During Preschool Ages] × [During School Ages] | | | -0.018 | | | -0.004 | | | -0.011* |
| | | | (0.014) | | | (0.004) | | | (0.006) |
| Sibling & Birth Cohort FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Subdistrict & Regency/City Controls | | Yes | Yes | | Yes | Yes | | Yes | Yes |
| Observations | 1,922 | 1,922 | 1,922 | 1,460 | 1,460 | 1,460 | 1,460 | 1,460 | 1,460 |
| Number of Groups (Siblings) | 1,491 | 1,491 | 1,491 | 1,176 | 1,176 | 1,176 | 1,176 | 1,176 | 1,176 |
| R-squared | 0.908 | 0.909 | 0.909 | 0.215 | 0.232 | 0.234 | 0.155 | 0.172 | 0.182 |

*Notes*: Standard errors are clustered at the subdistrict level in parentheses. Statistical significance is denoted as *** at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level. For brevity, this table presents only the coefficients of interest from Equation (2). The sample in this analysis consists of children aged 7 to 16 years old.

**Table 7:** Bias-Adjusted Effects on Health and Education: Accounting for Selection on Unobservables

| | Baseline Effect $\mathring{\beta}$, (S.E.), $[\mathring{R}]$ | Controlled Effect $\tilde{\beta}$, (S.E.), $[\tilde{R}]$ | Identified Set $[\tilde{\beta}, \beta^{*}]$ | Exclude Zero? | Within 95% CI? |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Panel A. Child Health | | | | | |
| Forest Loss (log) | | Dependent Variable: Fever | | | |
| During the Last 12 Months | 0.0390** | 0.0389** | [0.0385, 0.0389] | Yes | Yes |
| | (0.0180)[0.002] | (0.0178)[0.027] | | | |
| Between 1 and 2 Years Ago | -0.0209 | -0.0244 | [-0.0417, -0.0244] | Yes | Yes |
| | (0.0175)[0.002] | (0.0173)[0.027] | | | |
| Between 2 and 3 Years Ago | -0.0159 | -0.0151 | [-0.0151, -0.0121] | Yes | Yes |
| | (0.0163)[0.002] | (0.0159)[0.027] | | | |
| Panel B. Child Education | | | | | |
| Forest Loss (log) | | Dependent Variable: Years of Education | | | |
| During Preschool Age | -0.139*** | -0.129*** | [-0.129, -0.088] | Yes | Yes |
| | (0.034)[0.905] | (0.037)[0.909] | | | |
| During School Age | -0.067* | -0.057 | [-0.057, -0.013] | Yes | Yes |
| | (0.037)[0.905] | (0.044)[0.909] | | | |
| Interaction Term: [During Preschool Ages] × [During School Ages] | -0.027** | -0.018 | [-0.018, 0.020] | No | Yes |
| | (0.011)[0.905] | (0.014)[0.909] | | | |

*Notes*: Statistical significance is denoted as *** at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level. The controlled effect includes the full set of controls used in the regression analysis. In columns (1) and (2), clustered standard errors are reported in parentheses, and $R$-squared values are in brackets. The identified set in column (3) is bounded between $\tilde{\beta}$ and $\beta^{*}$ based on $R_{\max} = 3\tilde{R} - 2\mathring{R}$ and $\delta = 2$ using Equation (3). Column (4) indicates whether the identified set excludes zero, and column (5) indicates whether the bounds are within the 95% confidence interval of the controlled effects ($\tilde{\beta}$). Panel A presents the results for child health analysis, and Panel B gives the results for educational analysis.

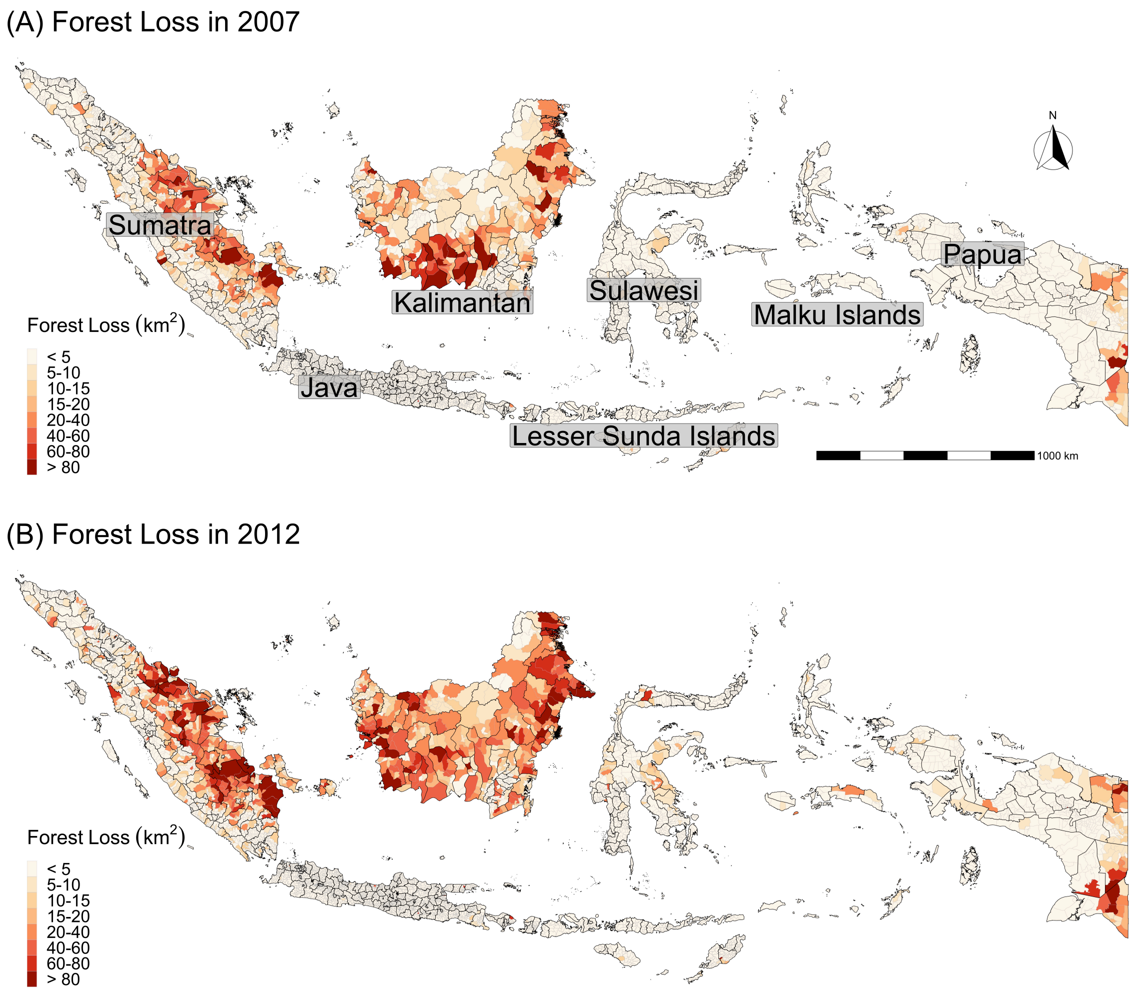


**Figure 1:** Forest Loss in Indonesia

*Notes*: Panel (A) shows the forest loss in 2007 and the names of islands, and Panel (B) presents the forest loss in 2012 at the subdistrict level.
*Source*: Authors' calculations using Hansen/UMD/Google/USGS/NASA.

# Online Appendix

## Appendix A Additional Tables

**Table A.1:** Summary Statistics: DHS Survey Months by Year (Main Sample)

| | Observation | |
|---|---|---|
| DHS Interview Month | 2007 | 2012 |
| 1 | - | 1 |
| 2 | - | - |
| 3 | - | 4 |
| 4 | - | 2 |
| 5 | - | 5,141 |
| 6 | - | 5,174 |
| 7 | 5,115 | 570 |
| 8 | 4,301 | 45 |
| 9 | 799 | - |
| 10 | 69 | - |
| 11 | 80 | - |
| 12 | 260 | - |
| Total | 10,624 | 10,937 |

**Table A.2:** Summary Statistics: Explanatory Variables

| | 2007 ($N$ = 10,624) | | 2012 ($N$ = 10,937) | |
|---|---|---|---|---|
| | Mean | S.D. | Mean | S.D. |
| *Child Characteristics* | | | | |
| Year of Birth | 2004.34 | 1.30 | 2009.18 | 1.31 |
| Age (Month) | 32.88 | 15.46 | 32.74 | 15.45 |
| Female | 0.47 | 0.50 | 0.48 | 0.50 |
| *Household and Parent Characteristics* | | | | |
| Rural | 0.68 | 0.47 | 0.61 | 0.49 |
| Poorest | 0.36 | 0.48 | 0.36 | 0.48 |
| Poorer | 0.22 | 0.41 | 0.21 | 0.41 |
| Middle | 0.17 | 0.38 | 0.18 | 0.39 |
| Richer | 0.14 | 0.35 | 0.14 | 0.35 |
| Richest | 0.10 | 0.31 | 0.10 | 0.30 |
| Female Headed HH | 0.05 | 0.22 | 0.07 | 0.25 |
| Firewood as Fuel | 0.57 | 0.50 | 0.43 | 0.49 |
| Private Toilet | 0.50 | 0.50 | 0.63 | 0.47 |
| Water on Premises | 0.69 | 0.46 | 0.46 | 0.50 |
| Piped Water | 0.16 | 0.36 | 0.10 | 0.30 |
| Water Treatment | 0.94 | 0.24 | 0.71 | 0.45 |
| N of Children Living Together | 2.62 | 1.59 | 2.49 | 1.56 |
| Father's Age | 39.27 | 11.58 | 40.83 | 12.42 |
| Mother's Age | 29.83 | 6.38 | 30.10 | 6.42 |
| Father's Years of Education | 8.54 | 4.06 | 9.18 | 4.18 |
| Mother's Years of Education | 8.16 | 3.93 | 9.03 | 4.22 |
| Agricultural Worker | 0.46 | 0.49 | 0.33 | 0.46 |
| Currently Working | 0.47 | 0.50 | 0.50 | 0.50 |
| *Region-Level Characteristics* | | | | |
| Survey Year's Forest Cover | 424,806 | 1,017,871 | 363,299 | 1,004,420 |
| Nighttime Lights | 6.93 | 9.09 | 10.98 | 13.84 |
| Precipitation (mm) | | | | |
| Last Month | 181.46 | 98.25 | 215.13 | 150.87 |
| Last 12 Months | 2,576.92 | 694.76 | 12,427.52 | 18,847.67 |
| Population Density | 428.50 | 1,077.81 | 617.30 | 1,484.61 |

*Notes*: This table reports the number of observations, means, and standard deviations (S.D.) of the explanatory variables used in the empirical analysis. This main sample consists of children aged 7 to 59 months old. The unit of the survey year's forest cover variable is one thousand square meters.

# Appendix B Data Description: Nighttime Lights, Precipitation, and Population

**Nighttime Lights Data**: In some cases, it has been recognized that economic development, such as urbanization and industrialization, accompanies deforestation. Furthermore, the incidence of child diseases may be associated with local economic activity. We use average nighttime lights to account for such economic factors that may be linked to deforestation and child health. The data come from the Version 4 DMSP-OLS Nighttime Lights Time Series from the National Geophysical Data Center at the National Oceanic and Atmospheric Administration (NOAA). The DMSP data were collected by the US Air Force Weather Agency and provided on a 30 arc-second grid (approximately 1 kilometer $\times$ 1 kilometer) of the GeoTIFF dataset. The average subdistrict-level nighttime lights during the deforestation period used in the analysis are created after adjustment for the gap in DHS survey months.

**Precipitation Data**: To account for climatic variations across subdistricts, this study uses precipitation data because, as discussed in Section 2.3, precipitation may be correlated with malaria development and transmission. Precipitation data are from the Climate Hazards Group Infrared Precipitation with Stations (CHIRPS). We use the global monthly information of CHIRPS version 2.0, which is gridded rainfall time series data at 0.05 $\times$ 0.05 degree spatial resolution (approximately 5 kilometers $\times$ 5 kilometers). The subdistrict-level average monthly precipitation is merged with our sample at the same level.

**Population Data**: Population and population density may also correlate with malaria infection because the intensity of malaria transmission relies on a human host. Thus, this paper uses information on population and area at the district level to control for population density. These variables come from the Indonesia Database for Policy and Economic Research (INDO-DAPOER).

# Appendix C Sensitivity Checks

## C.1 Sample with Different Age Groups

In this section, we run regressions similar to Table 3 with respect to different age groups. For the main analysis, the sample consists of children aged 7 to 59 months old because of their different immune systems, as discussed in Section 4.2. 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 of the effects on infants and young children focused on children under five years without considering differences in immune systems by age. Although, as discussed in Section 4.2, the age from zero to six months is considered one of the most important periods for child development, it is also suggested that the age of one year is an important turning point. Hence, we compare the results using the different age groups by estimating the same specification as in column (3) of Table 3.

Table C.1 reports these results: Column (1) presents the estimates for the sample of children aged 0 to 59 months old (under five years). The coefficient of forest loss for the last 12 months slightly decreases from 0.0389 in column (3) of Table 3 to 0.035 but remains statistically significant, and the coefficients of forest losses 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 3). Conversely, in column (3), we see no statistically significant impact of forest loss for the last 12 months for the infant sample, but there is a possible negative association between forest loss two and three years ago and fever incidence. As we theorized regarding the reason for the negative relationship between forest loss more than one year ago and fever incidence in Section 5.1, deforested areas may become an environment where mosquitoes cannot live in the long term, and this relationship was stronger for infants than for children under five years old. The findings in Table C.1 indicate that our main results are not sensitive to the selection of the age range, such as whether to include infants, and the incidence of infant fever is not affected by forest loss.

**Table C.1:** Sensitivity Checks: Sample with Different Age Groups

| Dependent Variable | Fever | | |
|---|---|---|---|
| | 0-59 Months Old | 13-59 Months Old | 0-6 Months Old |
| | (1) | (2) | (3) |
| Forest Loss (log) | | | |
| During the Last 12 Months | 0.0350** | 0.0385** | 0.0525 |
| | (0.0174) | (0.0186) | (0.0432) |
| Between 1 and 2 Years Ago | -0.0179 | -0.0267 | -0.0153 |
| | (0.0163) | (0.0180) | (0.0456) |
| Between 2 and 3 Years Ago | -0.0228 | -0.0181 | -0.0728* |
| | (0.0157) | (0.0167) | (0.0428) |
| 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 | 23,207 | 17,897 | 2,764 |
| Number of Subdistricts | 1,540 | 1,540 | 1,161 |
| R-squared | 0.028 | 0.024 | 0.135 |

*Notes*: Standard errors are clustered at the subdistrict level in parentheses. Statistical significance is denoted as ** at the 5 percent level and * at the 10 percent level. All columns include the log of the survey year's forest cover. For brevity, this table presents only the coefficients of interest from Equation (1). The sample in column (1) consists of children aged 0 to 59 months old, column (2) consists of children aged 13 to 59 months old, and column (3) consists of children aged 0 to 6 months old.

## C.2 Choice of Lag Variables of Forest Loss

How sensitive are our estimated results to the choice of lag variables of forest loss? Here, we see the results of another regression that includes the further lags of forest loss, three years before and four years before, in the main specification of column (3) of Table 3. The aim of this analysis is to determine whether our main results are sensitive to the choice of lag variables (i.e., whether the selection of lag variables was subjectively determined regarding the results, such as statistical significance).

Table C.2 gives the estimated results. As columns (1) and (2) indicate, adding these lag variables does not substantially change the estimates in Table 3, suggesting that the choice of our lag variables in the main analysis is not problematic. Concerns about the subjective choice of the lag variables are also ruled out. In addition, forest loss that occurred in the following year is included in this specification to confirm whether future forest loss is associated with the incidence of fever because future forest loss should not affect a child's current fever incidence. Column (3) adds forest loss during the following 12 months to column (2). The coefficients of forest loss for the last 12 months and one year before slightly increase in magnitude. However, the main pattern of the results is unchanged, and the coefficient of forest loss during the following 12 months is smaller and insignificant.

**Table C.2:** Sensitivity Checks: Choice of Lag Variables of Forest Loss

| Dependent Variable | Fever | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Forest Loss (log) | | | |
| During the Following 12 Months | | | -0.0121 |
| | | | (0.0142) |
| During the Last 12 Months | 0.0398** | 0.0412** | 0.0505** |
| | (0.0176) | (0.0179) | (0.0210) |
| Between 1 and 2 Years Ago | -0.0202 | -0.0252 | -0.0300 |
| | (0.0175) | (0.0185) | (0.0191) |
| Between 2 and 3 Years Ago | -0.0320 | -0.0297 | -0.0245 |
| | (0.0214) | (0.0216) | (0.0227) |
| Between 3 and 4 Years Ago | 0.0156 | 0.00961 | 0.00663 |
| | (0.0117) | (0.0125) | (0.0126) |
| Between 4 and 5 Years Ago | | 0.0101 | 0.00703 |
| | | (0.00871) | (0.00983) |
| 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 in parentheses. Statistical significance is denoted as ** at the 5 percent level. All columns include the log of the survey year's forest cover. For brevity, this table presents only the coefficients of interest from Equation (1). The sample in this analysis consists of children aged 7 to 59 months old.

# Appendix D Selection on Observables and Unobservables

In this section, we explain Oster's method used in Section 5.3 in more detail. This method centers on the estimation of bias-adjusted coefficients to assess the consistency of the estimates by accounting for the movements of both the coefficient and $R$-squared values when including observed controls in the baseline specification. In Table 7, we calculate the approximation of the bias-adjusted coefficient, $\beta^*$, using the following equation proposed by Oster (2019):

$$\beta^* \approx \tilde{\beta} - \delta[\mathring{\beta} - \tilde{\beta}]\frac{R_{\max} - \tilde{R}}{\tilde{R} - \mathring{R}}, \tag{3}$$

where $\tilde{\beta}$ and $\tilde{R}$ are the estimated coefficient and $R$-squared from a regression with the full set of observed controls, that is, the results from Equation (1). $\mathring{\beta}$ and $\mathring{R}$ are their equivalents from a baseline regression with only the relevant controls composing the identification assumption (i.e., we include only the subdistrict and survey year-month fixed effects).

To obtain $\beta^*$, we need to make assumptions regarding $\delta$ and $R_{\max}$. $\delta$ is defined as the ratio of the selection on unobservables to selection on observables.[33] We assume that $\delta = 2$, indicating that unobservables are related to the treatment status twice as much as observables.[34] $R_{\max}$ is the $R$-squared value in a regression when controlling for both observed and unobserved variables. Since it is an unknown parameter, we consider in a conservative way $R_{\max} = 3\tilde{R} - 2\mathring{R}$, assuming that the unobserved controls explain twice as much of the outcome as the observed controls. Thus, letting $\delta = 2$ and $R_{\max} = 3\tilde{R} - 2\mathring{R}$ together, we assume that the unobserved controls explain four times as much of the outcome as the observed controls. Coefficient stability is assessed according to whether the identified set, $[\tilde{\beta}$, $\beta^*(R_{\max}, \delta)]$, excludes zero and whether the bounds of the set are within the 95% confidence interval of $\tilde{\beta}$ (i.e., the controlled effect).

[33] Oster (2019) defined the proportional selection relationship as $\delta\frac{\sigma_{1X}}{\sigma_1^2} = \frac{\sigma_{2X}}{\sigma_2^2}$, where $\sigma_{iX}$ is the covariance of the variable of interest and observables and unobservables. See the discussion in Section 3.1 in Oster (2019) for details.

[34] We also use $\delta = 1$, assuming that observables and unobservables are equally correlated with forest loss, to estimate $\beta^*$. The results are nearly identical to the controlled effect in column (2) of Table 7; i.e., the calculated identified sets exclude zero and fall within the 95% confidence interval of the controlled effects.