When it Stings: Malaria and Student Achievement*

Tejendra P. Singh†

Georgia State University

September, 2024

Abstract

Climate change will expose previously unexposed geographies to vector-borne diseases. The effect of exposure to these diseases on human capital formation is crucial for understanding the costs associated with climate change. In this paper, I provide causal evidence on potential contemporaneous exposure to malaria and its impact on student achievement in Tanzania. I find that increased potential malaria exposure worsens student performance on high-stakes exams. These effects differ significantly across subjects and various sub-populations. I establish the robustness of my results through multiple empirical checks. Exploring potential mechanisms, it appears that the worsening of student exam performance may be driven by the illness of another household member due to potential exposure to malaria.

JEL Classifications: I10, I20, J24, O15, Q56

Keywords: Malaria, Test scores, Climate change, Tanzania

*I am indebted to Garth Heutel for constant encouragement and support throughout this project. Hee Kwon Seo provided incredible suggestions on longitudinal linking of the student achievement data. I thank Daniel Kreisman, Jim Marton, Jonathan Smith, and Olanrewaju Yusuff for many helpful comments and suggestions. Any remaining errors are my own.

†tpratapsingh1@student.gsu.edu

1 Introduction

In a recent report, the World Health Organization expressed concerns about the role of climate change in causing a surge in vector-borne diseases around the world. Climate change, coupled with other wider environmental changes like rapid urbanization, is spreading the risk of infection geographically further, even to previously unexposed regions.[1]

In addition to stretching the burdened healthcare system, this increased risk may also negatively impact other outcomes. For instance, malaria infection can lead to a reduction in productivity which may affect aggregate economic outcomes like poverty and economic growth (Hong, 2007). These effects could be more acute for low and lower-middle income countries, which are likely to be severely impacted due to heightened risk and exposure.

Given this backdrop, it is crucial to understand how exposure to vector-borne diseases affects human capital formation to fully quantify the costs of climate change. Furthermore, in some contexts, like mine, these costs may get accentuated due to regulatory frictions. For instance, in Tanzania, failure to appear for high-stakes exams leads to a costly and time consuming process of reappearing application. Existing work examining the effect of malaria exposure on human capital formation focuses on the long-term effects of early life malaria exposure (Barreca, 2010; Hong, 2007; Veras, 2022).

While malaria infection in early life severely hampers growth and development, contemporaneous malaria exposure may also affect human capital accumulation. For instance, students may be forced to be out of school due to malaria infection leading to worsening performance on tests (Cirera et al., 2022). Despite the apparent importance of the impact of contemporaneous malaria exposure on various outcomes, we do not have a good understanding of how these outcomes are impacted. In this paper, I attempt to fill this gap.

Using administrative data on student performance on high-stakes primary school exams in Tanzania, I causally identify the effect of potential malaria exposure in the days leading up to the exam on student performance on these exams. High-stakes nature of these exams emanates from their importance to educational attainment trajectory and, consequently, the labor market performance of students. To causally identify the effect of malaria exposure on student achievement, I rely on within student variation in potential malaria exposure on the eve of the exams achieved through longitudinal linking of student test scores data.

Additionally, my research design also controls for time-invariant characteristics of the subjects along with the exam-level and temporal shocks that are common across all examination centers in the country. I find that a one standard deviation increase in students' residence district malaria positivity rate worsens

[1] More information can be accessed on it here.

student performance by 0.029 standard deviations.

I establish the robustness of the results through multiple checks. I show that changing the analytical sample, changing the measure of potential malaria exposure, and estimating a more demanding empirical specification controlling for time-varying unobservable student characteristics do not alter the conclusions. I also do not find that the decline in academic achievement is driven by students selecting to appear for the exams.

While the heterogeneity analysis does not paint a consistent picture of student performance across subjects differing across various sub-populations, the main effect differs across students' sex, the economic performance of their residence region, and their age. I also do not find that the students are more likely to fail the exam due to potential exposure to higher levels of malaria infection in their geographically proximate population.

In examining potential mechanisms that may help explain the effect of contemporaneous malaria exposure on student performance, I use survey data from the National Panel Survey of Tanzania (NPS) 2014-15. Using these data, I show that the likelihood of a school-going household member reporting missing school due to the illness of another household member in the last two weeks is higher due to potentially higher levels of malaria exposure. Together with the result that the healthcare provider visits and hospitalization also increase due to potentially higher levels of malaria exposure, I show that the worsening of student performance on these high-stakes exams could be due to student school absence in the days leading up to exams. Consequently, the setbacks in education attainment due to contemporaneous exposure to malaria might be compounded in the long-run as household loses income due to healthcare expenditure leading to students being chronically absent from schools as they supplement household income by supplying their labor.

This paper contributes to a large body of work examining the effect of malaria exposure on human capital formation. One part of this strand of literature uses early life exposure to malaria on various outcomes (Barreca, 2010; Bleakley, 2010; Veras, 2022). The other part uses natural variation arising due to differences in malaria prevalence at the onset of malaria eradication campaigns in different countries (Barofsky et al., 2015; Bleakley, 2010; Cirera et al., 2022; Cutler et al., 2010; Klejnstrup et al., 2018; Kuecken et al., 2020; Lucas, 2010; Rawlings, 2016; Venkataramani, 2012). Almost unequivocally, these studies document negative effect of malaria exposure in early life on later life outcomes. I contribute to this literature by examining the effect of contemporaneous malaria exposure on student achievement. Finally, my work also contributes to the work on short-term effects of disease environment on human capital outcomes (Kuecken et al., 2020).

The rest of the paper proceeds as follows. Section 2 provides a brief description of the background of the setting. Section 3 discusses the data sources and presents descriptive statistics for the analytical sample. In Section 4, I discuss empirical strategy. Section 5 presents the results. Finally, Section 6 concludes.

2 Background

In this section, I provide a background of the setting. I begin with describing primary schooling in Tanzania.[2] Primary schooling in Tanzania is spread over seven years. Students can enroll in both government-owned and private schools. While primary school education has been free since 2001, the increased enrollment has led to a significant decline in the quality of education with an adverse effect on exam scores (Habyarimana et al., 2021).

Students in primary schools take two exams during their studies. The first exam takes place after fourth grade while the other takes place after seventh grade. The exam at the end of the fourth grade is called Standard Four National Assessment (SFNA). The primary school leaving exam at the end of the seventh grade is called Primary School Leaving Examination (PSLE).

Each exam at the SFNA level is 90 minutes long with at least four questions for each subject (NECTA, 2018). At the PSLE level, each exam consists of 45 questions. 40 of these questions are multiple-choice while five questions are open-ended response questions. Each exam at the PSLE level lasts for at least 90 minutes with some exams lasting for as long as two hours (NECTA, 2020).

Both SFNA and PSLE exams are high-stakes because students who miss or fail the exams can only appear again for these exams under special circumstances. Having passed the SFNA exams is a requirement for continuing primary schooling and to be eligible for registration for PSLE exams. Further, it is also the case that scores on PSLE exams determine the education trajectory of students. PSLE performance determines the eligibility of the student to enroll in secondary school.

I next turn to discuss malaria in Tanzania. Malaria affects a large part of Tanzania both in terms of population and area at risk of infection. In addition to imposing mortality and morbidity burden, malaria also negatively impacts human capital accumulation and economic productivity (Mboera et al., 2007). While there have been multiple initiatives to reduce malaria risk, eradication of disease remains a challenge partly due to failure of compliance with existing malaria control strategies (Douglas et al., 2021).

While the malaria infection persists throughout the year, the risk is elevated during rainfall season (Oesterholt et al., 2006). This is due to climatic conditions that render a large part of the country a fertile space for the breeding of mosquitoes responsible for malaria transmission. Malaria is a vector-borne disease primarily transmitted through female Anopheles mosquitoes.[3]

[2] Readers interested in learning more about this can see Mtenga and Singh (2022) and the references therein.
[3] More information on malaria can be accessed at www.cdc.gov/malaria/about/biology/index.html.

3 Data and Descriptive Statistics

An ideal data to uncover the causal effect of contemporaneous malaria exposure on student achievement would contain information on the malaria infection status of the student, their academic performance, and other correlates of the two. While I am unable to access such data, I combine multiple data to construct the analytical sample. I describe each data along with its source in this section. I also discuss descriptive statistics from the analytical sample.

3.1 Student Test Score Data

I use student test score data from Tanzania. Data on academic achievement for two primary school exams – Standard Four National Assessment (SFNA) and Primary School Leaving Examination (PSLE) – is used. Both these exams are conducted by the National Examination Council of Tanzania (NECTA). NECTA is a national government agency in Tanzania. NECTA is responsible for conducting various national-level exams, including school exit exams. I provide more information on primary schooling in Tanzania and two exams in Section 2.

I scrape the universe of student test scores for SFNA exams conducted in the years 2015 and 2016 and PSLE exams conducted in 2018 and 2019. These data provide information on examination centers where students appeared for the exams. I also see what score students received for each exam they enrolled for. All students taking SFNA and PSLE appear for exams in English, Kiswahili, Mathematics, Science, and Social Knowledge. Additionally, I can see if the student was absent for a particular exam that they registered for. In these data, I also have information on the average score of the student. I also see the sex of the student. I do not have information on any other demographic of the student except sex.

The grading in both SFNA and PSLE is based on letter grades. I construct a numeric point measure corresponding to these letter grades. Grade "A" is assigned four points, Grade "B" three points, Grade "C" two points, Grade "D" one point, and Grade "E" gets zero points. I also construct a measure of aggregate score across all exams that the student takes at a given exam-level.

I longitudinally link students across SFNA and PSLE exams. This linking is done conditional on the examination center and the sex of the student. A match is determined based on the string similarity of the student name. For the analytical sample, I keep all students who have a Levenshtein distance of at least 0.75 and have a unique potential match based on these criteria. In a robustness check, I show that restricting the analytical sample only to those students who have a complete string match across two exam levels leads to almost identical point estimates without a significant loss to statistical precision.

3.2 Malaria Exposure Data

I use data on malaria exposure from the Tanzania Ministry of Health's National Portal (NHP). I obtain information on the malaria positivity rate among pregnant women during their first antenatal care (ANC) visit (Thawer et al., 2020). NHP provides this number as a percentage of positive malaria tests among pregnant women during their first ANC visit. For testing, NHP reports using either a rapid diagnostic test (RDT) or blood slides.

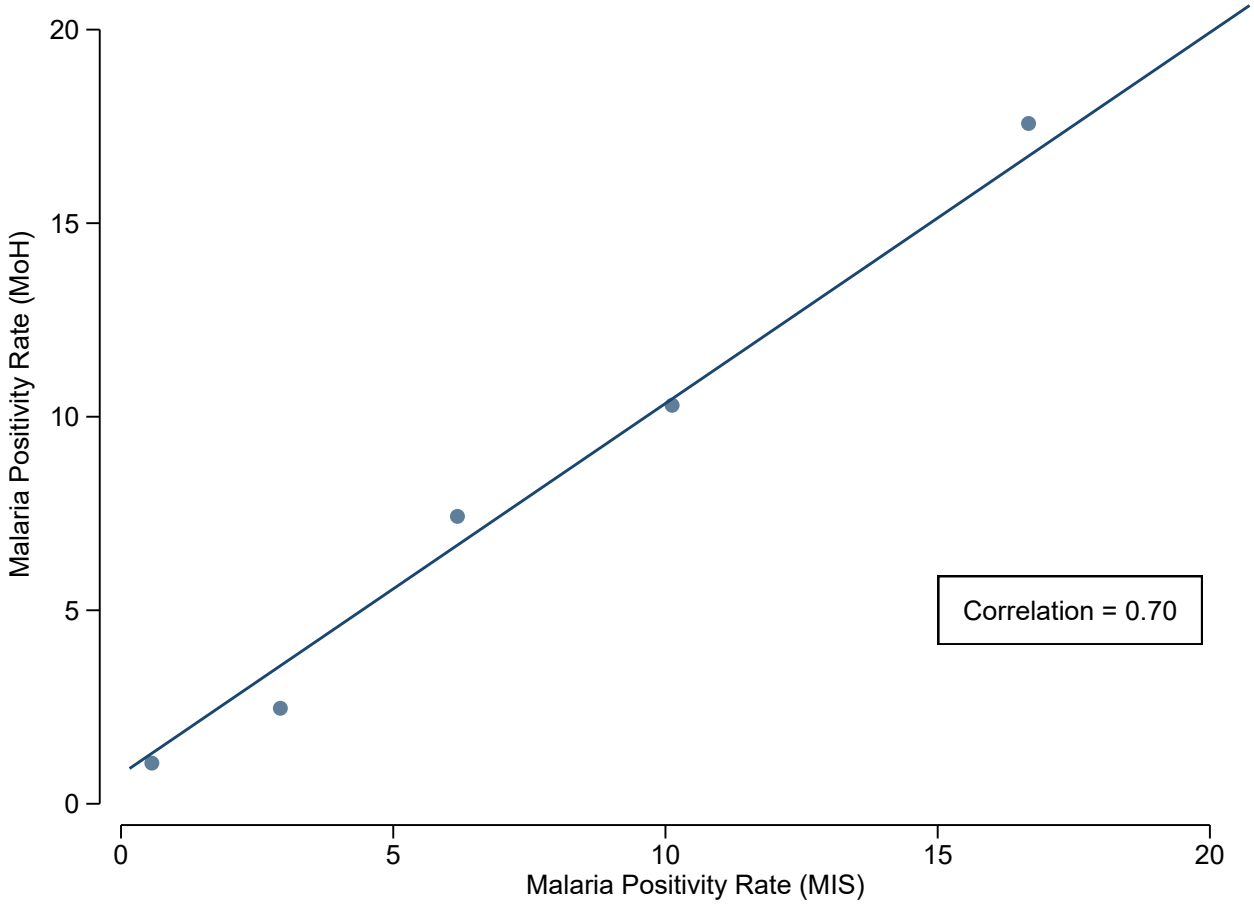
Data from NHP is scraped for all districts and months for the sample period of 2016 to 2019. While NHP data is the most accurate publicly available data on malaria exposure, it has a few limitations. First, I can't verify if the positivity rate among pregnant women corresponds to the positivity rate in the wider population. This is due to unequal access to healthcare and the potential migration of pregnant mothers from their usual place of residence. Second, there are potential non-random missing observations in the administrative data that we cannot control for in our empirical specifications.[4]

As I am unable to obtain information on whether the student is infected with malaria or not, I assume that the positivity rate among pregnant mothers is reflective of the extent of malaria infection amongst the students. How reasonable is this assumption? I use data from the Tanzania Demographic and Health Survey-Malaria Indicator Survey (MIS), 2015–2016. For children between the ages of six to fifty-nine months, MIS provides information if the child tests positive for malaria using rapid diagnostic testing. Figure 1 presents a binned scatter plot and regression estimate for the malaria positivity measure at the regional level in MIS and NHP data. While not perfectly correlated, the two malaria positivity rate measures are highly correlated. I emphasize that the two measures reflect malaria positivity rates in two distinct populations. Despite the difference in the underlying population, the high correlation between the two data sources indicates that it is not unreasonable to assume that the malaria positivity rate among pregnant mothers in NHP data is a good proxy for malaria infection in the student population.

I provide descriptive statistics using NHP data in Figure 2 and Figure 3. Figure 2 provides an empirical kernel density estimate of the malaria positivity rate for the analytical sample. I confirm the unimodal distribution of the positivity rate and leverage the variation depicted in the figure in the empirical specifications. In Figure 3, I show how the malaria positivity rate evolves during our sample period. Over the years, there has been a consistent decline in the malaria positivity rate. I also note that there is wide variation across months in malaria positivity rate for each year in the analytical sample.

[4]There are less than 1.3% missing values for malaria positivity rate data from NHP.

Figure 1: Correlation between Malaria Positivity Rate in MIS and NHP

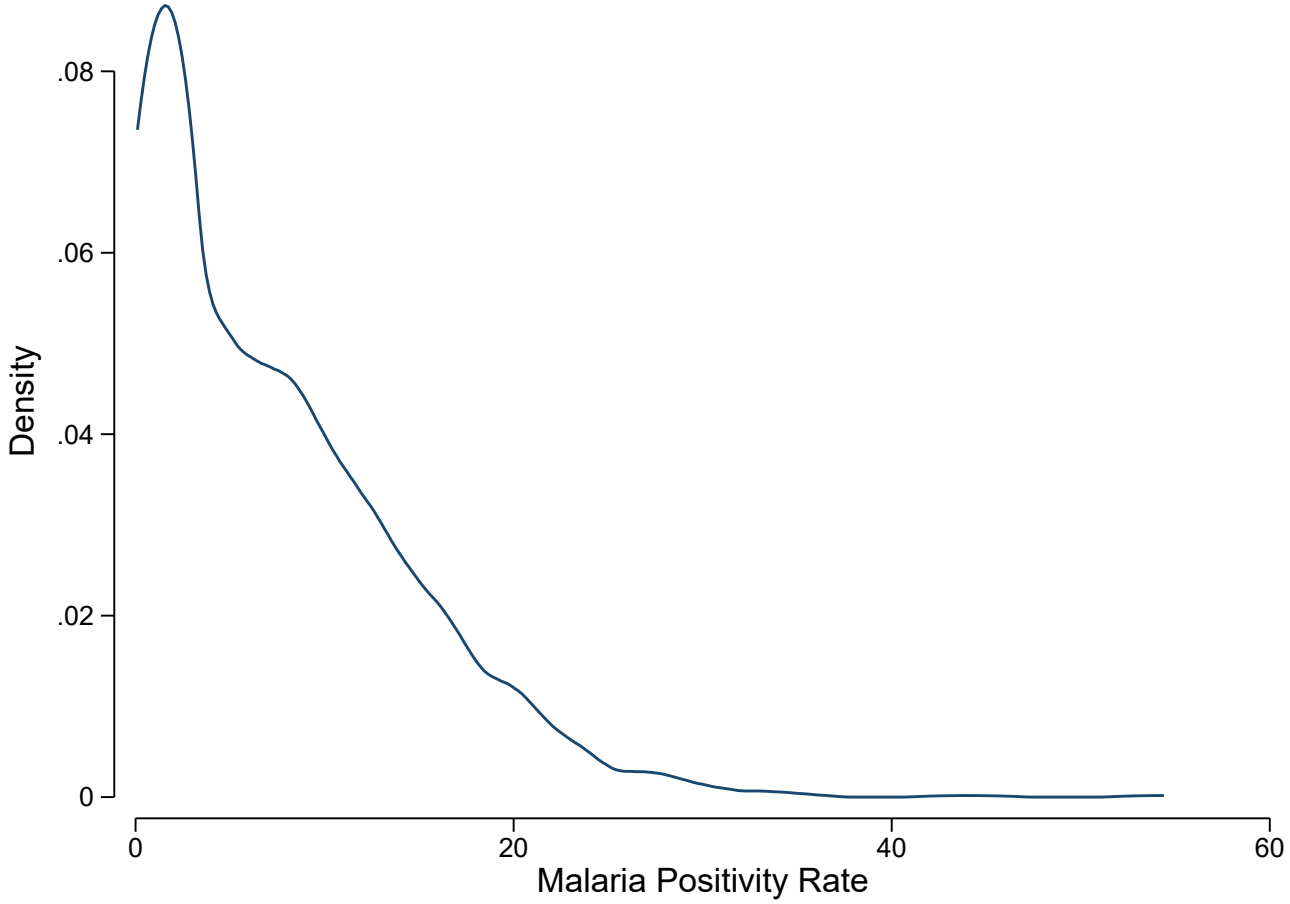


Note: This figure plots binned scatter and least square estimates for Tanzania Demographic and Health Survey-Malaria Indicator Survey (MIS), 2015–2016, and Tanzania Ministry of Health's National Portal (NHP) malaria positivity rate. MIS malaria positivity rate measure is among the children in the surveyed households who are between the ages of six to fifty-nine months. NHP malaria positivity rate measure is the percentage of positive malaria tests among pregnant women during their first ANC visit. Data from MIS is aggregated up to month-year and region level using the survey weights to account for complex survey design. Data for region-month pairs in the MIS data is used (October to December 2017). The figure plots the binscatter least square estimates derived using the methods in Cattaneo et al. (2023). The degree of global polynomial regression is set to one. The correlation coefficient between the two data series is shown.

3.3 Other Data

I use ERA5-Land monthly data on weather conditions to construct measures of weather conditions (Hersbach et al., 2020). I use information on temperature and precipitation from these data. These data are available in a gridded format with a spatial resolution of approximately 10×10 km. Satellite reanalysis data are being increasingly used in economics research and are shown to perform well in various climatic conditions (Donaldson and Storeygard, 2016; Jain, 2020; Mordecai et al., 2020).

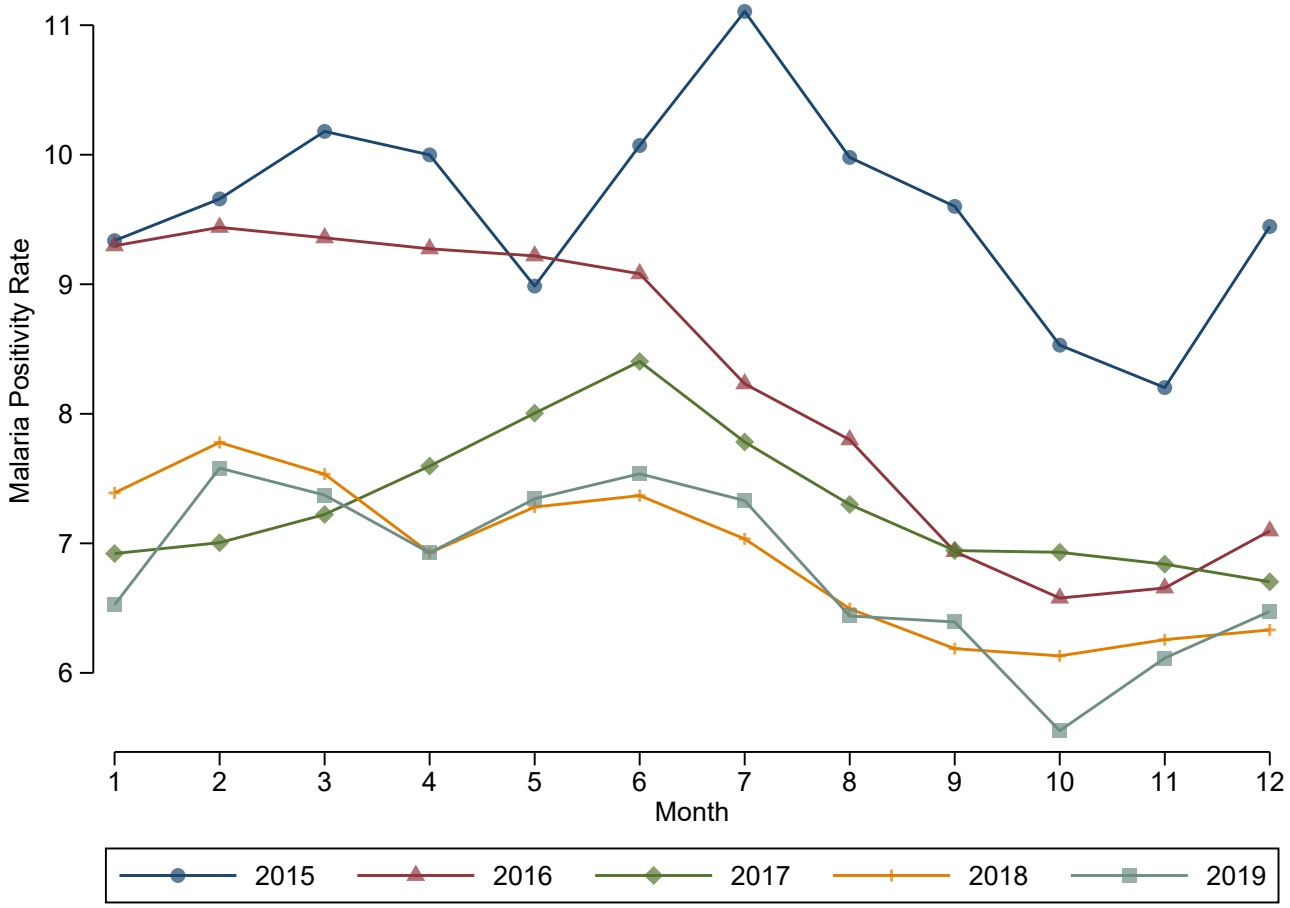
Figure 2: Malaria Positivity Rate Empirical Density



Note: Malaria positivity rate data comes from the Tanzania Ministry of Health's National Portal (NHP). These data are malaria positivity rate among pregnant women during their first antenatal care (ANC) visit (Thawer et al., 2020). All districts that appear in NHP are used. Data for month-year combinations in the analytical sample are used (August and November 2016 to 2019). Epanechnikov kernel density function with "optimal" bandwidth is used.

To examine the mechanisms that may lead to observed main effects, I use data from the National Panel Survey (NPS) of 2014/2015 implemented by the National Bureau of Statistics (NBS) of Tanzania. I refer to these data as NPS 2014 henceforth. NPS 2014 is a nationally representative household survey which is collected by the National Bureau of Statistics (NBS) of Tanzania. NPS 2014 is part of the Living Standards Measurement Study–Integrated Surveys on Agriculture (LSMS-ISA) Initiative supported by the World Bank. For my estimations, I use the fourth round of NPS, which was conducted between October 2014 and January 2016. NPS 2014 follows a two-stage sampling process where in the first stage, survey clusters were randomly selected for each rural and urban stratum of each region in Tanzania. In the second stage, households are randomly selected for each survey cluster.

Figure 3: Malaria Positivity Rate



Note: Malaria positivity rate data comes from the Tanzania Ministry of Health's National Portal (NHP). These data are malaria positivity rate among pregnant women during their first antenatal care (ANC) visit (Thawer et al., 2020). All districts that appear in NHP are used. The monthly average for each year is constructed by taking the arithmetic mean of the malaria positivity rate across all districts.

NPS 2014 contains detailed information on the agricultural activities of households, non-farm enterprises owned by household members, socioeconomic characteristics, consumption, and other expenditures. In my analysis, I rely on data about the schooling of household members along with information on their health and healthcare utilization. I discuss in detail the empirical strategy and variables used for examining the potential mechanisms in Section 4.

3.4 Descriptive Statistics

In addition to descriptive statistics presented in Section 3.2, I also present summary statistics for the analytical sample in Table 1. For test score data, I highlight variations in both aggregate and standardized

test scores. I also present summary statistics for malaria exposure data in the table. For malaria exposure, I notice a substantial variation in the analytical sample.

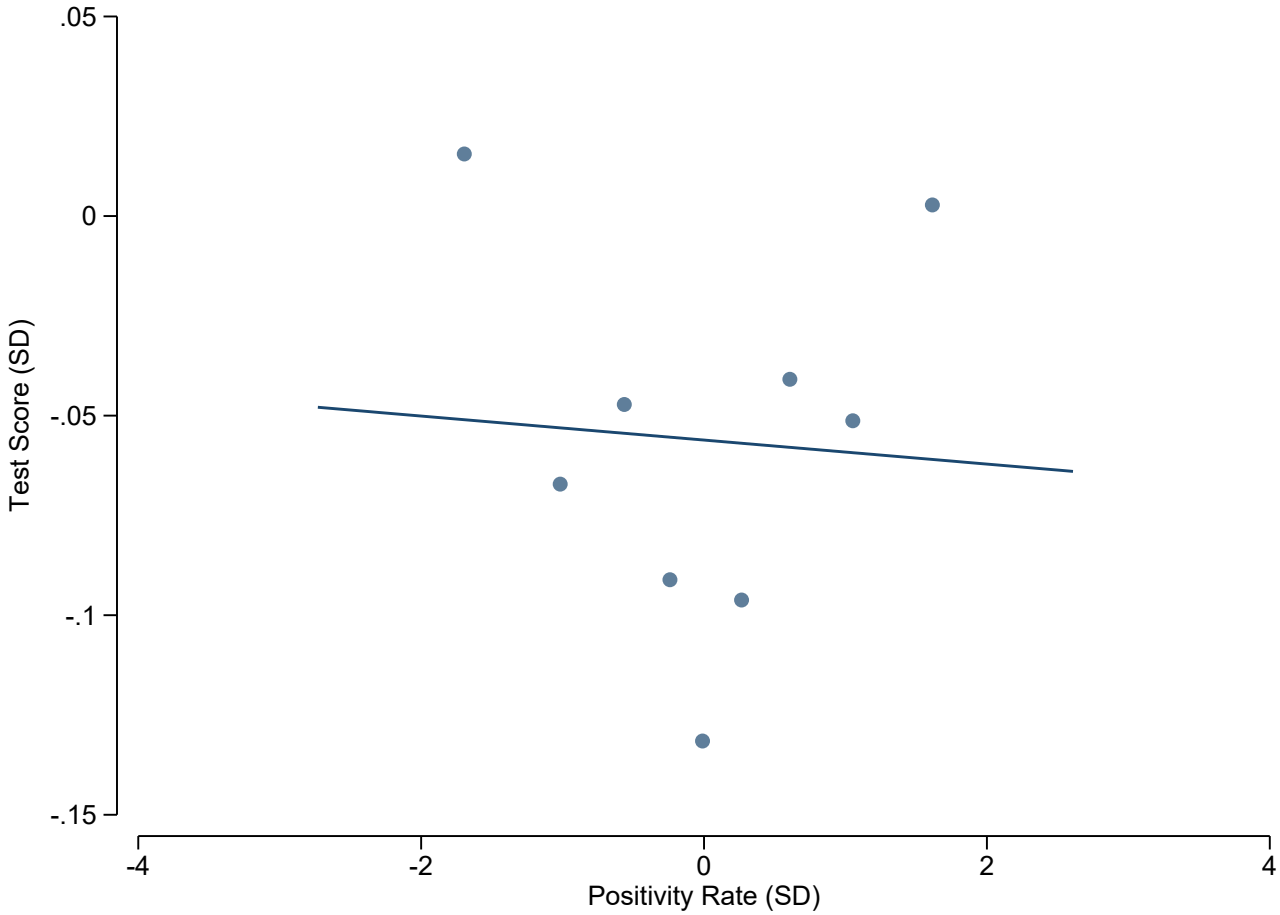
Table 1: Summary Statistics

	N	Mean	SD	Min	Max
SFNA Exams					
Total Score	1,237,859	11.669	5.113	0.00	28.00
Average Score	1,219,834	1.987	0.838	0.00	4.00
English	1,231,133	1.358	1.132	0.00	4.00
Kiswahili	1,231,229	2.787	1.085	0.00	4.00
Mathematics	1,231,119	1.440	1.080	0.00	4.00
Science	1,231,220	2.049	0.966	0.00	4.00
Social Knowledge	1,230,927	1.748	0.913	0.00	4.00
Career, Skills, Personality, and Sports	1,231,217	1.978	0.959	0.00	4.00
Failed	8,931,984	0.107	0.310	0.00	1.00
Absent	10,186,852	0.001	0.023	0.00	1.00
PSLE Exams					
Total Score	1,237,859	10.255	4.231	0.00	20.00
Average Score	1,221,297	2.126	0.858	0.00	4.00
English	1,221,202	1.686	1.008	0.00	4.00
Kiswahili	1,221,212	2.558	0.995	0.00	4.00
Mathematics	1,221,223	1.971	1.126	0.00	4.00
Science	1,221,227	2.167	0.896	0.00	4.00
Social Knowledge	1,221,248	2.012	0.858	0.00	4.00
Failed	7,420,416	0.057	0.232	0.00	1.00
Absent	8,659,845	0.000	0.000	0.00	0.00
Independent Variable					
Malaria Positivity Rate	17,739,657	6.579	6.498	0.10	34.00

Notes: Each observation in all rows corresponds to a student-subject pair. Malaria positivity rate data comes from the Tanzania Ministry of Health's National Portal (NHP). These data are malaria positivity rate among pregnant women during their first antenatal care (ANC) visit (Thawer et al., 2020). The total score is the sum of exam scores in each subject that the student takes the exam. Row labels denote the subject. Letter grades provided by the National Examination Council of Tanzania (NECTA) are converted into numeric points with letter grade "A" receiving four numeric points, letter grade "B" receiving three points, letter grade "C" receiving two points, letter grade "D" receiving one point, and letter grade "E" receiving zero point. The sample contains data from the Standard Four National Assessment (SFNA) exams conducted in 2015 and 2016 and the Primary School Leaving Examination (PSLE) exams conducted in 2018 and 2019. The candidates who appear for SFNA exams conducted in 2015 are longitudinally linked to PSLE exams conducted in 2018 and those who appear for SFNA exams conducted in 2016 are linked to PSLE exams conducted in 2019. Number of observations changes across the columns as not all students appear for all exams and for some students, NECTA does not report average letter grades.

I show the raw correlation between standardized test scores and malaria positivity rate in Figure 4. This

Figure 4: Test Score vs. Malaria Positivity Rate



Note: This figure plots binned scatter and least square estimates for standardized exam scores and malaria positivity rate. Malaria positivity rate data comes from the Tanzania Ministry of Health's National Portal (NHP). These data are malaria positivity rate among pregnant women during their first antenatal care (ANC) visit (Thawer et al., 2020). Standardized exam score for each subject is calculated over the sample of students in the corresponding exam-year and subject pair. Letter grades provided by the National Examination Council of Tanzania (NECTA) are converted into numeric points with letter grade "A" receiving four numeric points, letter grade "B" receiving three points, letter grade "C" receiving two points, letter grade "D" receiving one point, and letter grade "E" receiving zero point. The sample contains data from the Standard Four National Assessment (SFNA) exams conducted in 2015 and 2016 and the Primary School Leaving Examination (PSLE) exams conducted in 2018 and 2019. The candidates who appear for SFNA exams conducted in 2015 are longitudinally linked to PSLE exams conducted in 2018 and those who appear for SFNA exams conducted in 2016 are linked to PSLE exams conducted in 2019. The figure plots the binscatter least square estimates derived using the methods in Cattaneo et al. (2023). The degree of global polynomial regression is set to one.

figure suggests that higher levels of malaria exposure in the days leading up to the exam month have a detrimental impact on student performance. I, however, stop short of identifying this as a causal relationship due to the absence of various controls correlated with both malaria positivity rate and test scores. To uncover causal estimates, I discuss empirical strategy next.

4 Empirical Strategy

This study aims to uncover the causal effect of contemporaneous malaria exposure on student achievement. My main empirical specification to this end employs student fixed-effects. I estimate the following specification using ordinary least squares.

$$y_{i,s\times l} = \alpha_i + \alpha_{s\times l} + \alpha_{\{s\times l\}(y)} + \beta MalariaPositivityRate_{i(d),\{s\times l\}(m)} + \varepsilon_{i,s\times l} \tag{1}$$

In Equation 1, $y_{i,s\times l}$ is the score for student i on the exam for subject s and exam-level l for which the student appears. This specification controls for student fixed-effects, denoted by α_i. These fixed-effects control for time-invariant student characteristics that are correlated both with the exposure to malaria on the eve of exams and the performance on these exams. Due to controlling for these fixed-effects, I leverage variation within a student in the exposure to malaria on the eve of exams to estimate the effect on student achievement. I am able to leverage this variation as each student in my analytical sample appears for two different exam-levels three years apart. Equation 1 also includes subject and exam-level (either Standard Four National Assessment (SFNA) or Primary School Leaving Examination (PSLE)), fixed-effects. These are denoted by $\alpha_{s\times l}$ in Equation 1. These fixed-effects control for subject and exam-level unobservable characteristics correlated with the outcomes and malaria exposure measures such as advanced content in exams for higher grades. Failure to account for subject and exam-level fixed-effects risk attributing change in student performance to these unobservables rather than the extent of malaria exposure on the eve of exams. When the outcome variable is performance on a specific subject, I replace subject and exam-level fixed-effects with fixed-effects specific to each exam-level.

Equation 1 also controls for temporal shocks that are common across the students by including exam-year fixed-effects. There are denoted by $\alpha_{\{s\times l\}(y)}$. These fixed-effects account for annual changes that are experienced by all the students. For instance, the budget outlays for the U.S. President's Malaria Initiative have increased annually during the sample period (Barat et al., 2023). I risk attributing changes in the outcomes to these changes if I do not account for the exam-year fixed-effects. $\varepsilon_{i,s\times l}$ are idiosyncratic errors that I cluster at the examination center (or school) level.

The independent variable of interest in Equation 1 is $MalariaPositivityRate_{i(d),\{s\times l\}(m)}$. This is the malaria positivity rate in the district of residence of student i, $i(d)$, in the month during which the student appears for the exam in subject s and exam-level l, $\{s\times l\}(m)$. I note that for SFNA exams I use the malaria positivity rate in November of the year in which the student appears for SFNA exams and August for PSLE exams. Further, the malaria positivity rate in the district of residence of the student corresponds to the malaria positivity in that district among pregnant women during their first antenatal care (ANC) visit. The parameter of interest, β, is the marginal effect of one unit increase in the malaria positivity rate on the outcome variable.

In order for the estimate of β to be interpreted causally, it must be true that the errors $\varepsilon_{i,s\times l}$ are uncorrelated with the malaria exposure after accounting for the fixed-effects. This will be the case if there are student-level time-varying characteristics that I cannot observe. In order to establish that my estimates are not plagued by these time-varying unobservables, in a robustness check I show that the main estimates are unaltered by replacing student fixed-effects in Equation 1 with student-level time-trends. As I am unable to ascertain if a student is infected with malaria, the estimate of β is an intention-to-treat estimate of malaria exposure on student achievement.

I next turn to discuss specifications that I implement in order to examine mechanisms that may explain my main effects. As noted in section 3, I rely on National Panel Survey (NPS) 2014 data to estimate these specifications. The specification that I estimate is as follows.

$$y_i = \nu malaria_{d(m(i))} + \mathbf{X}_{d(i)}\xi + \mathbf{X}_i\pi + \mathbf{X}_{h(i)}\rho + \sigma_i \tag{2}$$

In equation (2), $malaria_{d(m(i))}$ is malaria positivity rate in district d where respondent i reside in the month of interview m. In this specification, I also include district-level controls, individual controls, and household controls denoted by $\mathbf{X}_{d(i)}$, $\mathbf{X}_i$, and $\mathbf{X}_{h(i)}$, respectively. The vector of district-level controls contains total precipitation in the month of interview m and its square. The vector of individual controls contains the sex of the respondent and their age. The vector of household controls consists of the number of members in the household, sex of the household head, total household expenditure on education in the last 12 months, an indicator for whether the household has access to electricity, total household expenditure on the consumption of food in last week, a household asset index, and an indicator for whether the household owns a non-farm enterprise or not.

I construct the household asset index as the sum of indicators for owning a bicycle, motorcycle, car/other vehicles (vans), tractor, computer, telephone, cellular, radio, television, refrigerator, and stove. Therefore, this index ranges from 0 to 11. Total household expenditure on education in the last 12 months is constructed by aggregating such consumption for all household members.

As I leverage only cross-sectional variation in Equation 2, I do not interpret the estimated effect of malaria exposure on the outcomes as causal. The estimated effects are suggestive and help points to the potential mechanisms that might be causing changes in student performance due to malaria exposure. Apriori it is unclear what the direction of bias would be.

5 Results

5.1 Main Results and Robustness Checks

Table 2 presents estimates from Equation 1. The first column of the table presents estimates of the specification where the dependent variable is the total score across all the exams that the student appears at a given exam-level. The point estimate reflects that the total score of the students who reside in districts with higher malaria positivity rates decreases. For each standard deviation (SD) increase in malaria positivity rate, there is a decline in total score of 0.029 SD. In the second column of the table, I pool scores in all subjects that the student takes exam at different exam-levels. The point estimates show that potential exposure to elevated malaria levels worsens student achievement across all subjects that students are tested on. This conclusion is strengthened with point estimates for all individual subjects except Science (which I return to later in this subsection) being negative and statistically significant as shown in column (4) to column (8) of Table 2. Point estimates for reading (English) subjects experience a relatively more pronounced worsening of performance than for quantitative (mathematics) subjects (p-value < 0.001).

Table 2: Effect of Malaria Exposure on Academic Achievement

	Total Score	All Subjects	Average Score	English	Kiswahili	Mathematics	Science	Social Knowledge
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Positivity Rate (SD)	-0.029***	-0.021***	-0.024***	-0.036***	-0.030***	-0.031***	0.016**	-0.028***
	(0.008)	(0.006)	(0.007)	(0.008)	(0.005)	(0.007)	(0.008)	(0.007)
Student FE	✓	✓	✓	✓	✓	✓	✓	✓
Exam-level FE	✓		✓	✓	✓	✓	✓	✓
Exam-year FE	✓	✓	✓	✓	✓	✓	✓	✓
Subject × Exam-level FE		✓						
Adj. R^2	0.521	0.495	0.472	0.426	0.482	0.458	0.321	0.337
N	2,332,647	15,240,211	2,299,700	2,310,316	2,310,416	2,310,321	2,310,424	2,310,166

Notes: Heteroskedasticity robust standard errors clustered by the school (examination center) are in parentheses. (* p<.10 ** p<.05 *** p<.01). Each observation in all columns corresponds to a student-subject pair. The Independent variable of interest in each column is the standardized malaria positivity rate. Standardized malaria positivity rate is calculated over all sample years and districts. Malaria positivity rate data comes from the Tanzania Ministry of Health's National Portal (NHP). These data are malaria positivity rate among pregnant women during their first antenatal care (ANC) visit (Thawer et al., 2020). The dependent variable in all columns is the standardized exam score for the subject. The column header denotes the subject. The total score in the first column is the sum of exam scores in each subject that the student takes the exam. Standardized exam score for each subject is calculated over the sample of students in the corresponding exam-year and subject pair. Letter grades provided by the National Examination Council of Tanzania (NECTA) are converted into numeric points with letter grade "A" receiving four numeric points, letter grade "B" receiving three points, letter grade "C" receiving two points, letter grade "D" receiving one point, and letter grade "E" receiving zero point. Except second column each specification in all columns includes student, exam-level, and exam-year fixed-effects. Exam-level fixed-effects in the second column are replaced with a fixed-effect specific to each subject and exam-level pair. The sample contains data from the Standard Four National Assessment (SFNA) exams conducted in 2015 and 2016 and the Primary School Leaving Examination (PSLE) exams conducted in 2018 and 2019. The candidates who appear for SFNA exams conducted in 2015 are longitudinally linked to PSLE exams conducted in 2018 and those who appear for SFNA exams conducted in 2016 are linked to PSLE exams conducted in 2019. Number of observations changes across the columns as not all students appear for all exams and for some students, NECTA does not report average letter grades.

I next establish the robustness of worsened performance due to potential exposure to higher levels of malaria infection in Table 3. In Panel A of the table, I repeat point estimates presented in Table 2. Panel B of the table modifies the construction of malaria infection exposure. As PSLE exams happen in the first week of September, my baseline specifications use malaria positivity rates in August as the measure of potential malaria exposure in the days leading up to PSLE exams. In Panel B, I use an average of malaria positivity rates in students' residence district in August and September as the measure of potential malaria exposure for students who appear for PSLE exams. Potential malaria exposure for SFNA exams continues to be the malaria positivity rate in November of the relevant exam-year. The point estimates in this panel show that our conclusions of worsening student academic performance due to potential exposure to higher levels of malaria infection are not driven by the specific way in which I construct potential malaria exposure of students.

In Panel C of Table 3, I restrict the analytical sample to only those students who have a complete string match for their names across SFNA and PSLE exams. This additional restriction on the analytical sample leads to a decline in the sample size. Point estimates from this restricted sample, however, are virtually the same as the point estimates for the baseline sample in which all students who have string similarity scores of at least 0.75 across SFNA and PSLE exams constitute the analytical sample.

Panel D of Table 3 presents point estimates from a placebo check. In this panel, I replace the malaria positivity rate to the month following the exams. Thus, for relevant exam-years malaria positivity rate in October in the students' residence district is used as a measure of potential malaria infection exposure for PSLE exams, and the malaria positivity rate in the district in December is used as the measure of potential malaria infection exposure for SFNA exams. If the malaria positivity rates in these months influence student achievement, then my estimates are conflated by unaccounted factors. Reassuringly, I do not find that the malaria positivity rate in the month following the conclusion of the exams affects student performance in these exams. This result also bolsters my claim of contemporaneous potential exposure to malaria influencing student performance on high-stakes primary school exams.

In order to establish that my point estimates do not suffer from bias arising due to time-varying unobservable student characteristics, Panel E of Table 3 presents point estimates from modifying Equation 1. I replace student fixed-effects in Equation 1 with student time-trends in this panel. Point estimates from these specifications are very close to the baseline point estimates in Panel A of Table 3. This suggests that time-varying student specific unobservables are unlikely to bias my estimates of the influence of potential malaria exposure on student achievement. Furthermore, the absence of statistical significance on Science exam scores demonstrates that improved student performance on Science exams in Table 2 could be driven by time-varying unobservables that my empirical framework cannot fully account for. Going forward, I continue to present estimates for Science examinations but not interpret them.

Table 3: Robustness Checks

	Absent	Total Score	All Subjects	Average Score	English	Kiswahili	Mathematics	Science	Social Knowledge
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Baseline									
Positivity Rate (SD)		-0.029***	-0.021***	-0.024***	-0.036***	-0.030***	-0.031***	0.016**	-0.028***
		(0.008)	(0.006)	(0.007)	(0.008)	(0.005)	(0.007)	(0.008)	(0.007)
Adj. R^2		0.521	0.495	0.472	0.426	0.482	0.458	0.321	0.337
N		2,332,647	15,240,211	2,299,700	2,310,316	2,310,416	2,310,321	2,310,424	2,310,166
Panel B: Alternate Malaria Exposure Measure									
Positivity Rate (SD)		-0.020**	-0.017**	-0.020***	-0.034***	-0.034***	-0.025***	0.025***	-0.025***
		(0.008)	(0.007)	(0.008)	(0.009)	(0.006)	(0.007)	(0.008)	(0.008)
Adj. R^2		0.524	0.496	0.475	0.433	0.483	0.461	0.323	0.338
N		2,363,178	15,422,027	2,330,003	2,340,619	2,340,718	2,340,623	2,340,727	2,340,469
Panel C: Complete String Match									
Positivity Rate (SD)		-0.028***	-0.021***	-0.022***	-0.033***	-0.031***	-0.030***	0.016*	-0.023***
		(0.008)	(0.007)	(0.008)	(0.009)	(0.006)	(0.007)	(0.008)	(0.008)
Adj. R^2		0.520	0.496	0.470	0.427	0.482	0.457	0.317	0.333
N		1,234,008	8,063,849	1,217,724	1,223,098	1,223,137	1,223,093	1,223,161	1,223,038
Panel D: Placebo Check									
Positivity Rate (SD)		0.039	0.026	0.026	0.014	-0.020	0.028	0.062*	0.019
		(0.036)	(0.029)	(0.033)	(0.034)	(0.022)	(0.029)	(0.032)	(0.030)
Student FE		✓	✓	✓	✓	✓	✓	✓	✓
Exam-level FE		✓		✓	✓	✓	✓	✓	✓
Exam-year FE		✓	✓	✓	✓	✓	✓	✓	✓
Subject × Exam-level FE			✓						
Adj. R^2		0.528	0.496	0.478	0.438	0.487	0.464	0.326	0.339
N		2,375,675	15,536,144	2,342,179	2,353,150	2,353,258	2,353,154	2,353,275	2,352,990
Panel E: Add Student Time-trends									
Positivity Rate (SD)		-0.029***	-0.022***	-0.023**	-0.037***	-0.030***	-0.031***	0.014	-0.028***
		(0.010)	(0.007)	(0.010)	(0.011)	(0.007)	(0.010)	(0.011)	(0.010)
Student Time-trends		✓	✓	✓	✓	✓	✓	✓	✓
Exam-level FE		✓		✓	✓	✓	✓	✓	✓
Subject × Exam-level FE			✓						
Adj. R^2		0.521	0.495	0.471	0.426	0.482	0.458	0.321	0.337
N		2,332,647	15,240,211	2,299,700	2,310,316	2,310,416	2,310,321	2,310,424	2,310,166
Panel F: Student Absent									
Positivity Rate (SD)	-0.00005								
	(0.00004)								
Student FE	✓								
Exam-level FE	✓								
Exam-year FE	✓								
Adj. R^2	0.058								
N	15,399,578								

Notes: Heteroskedasticity robust standard errors clustered by the school (examination center) are in parentheses. (* p<.10 ** p<.05 *** p<.01). Each observation in all columns corresponds to a student-subject pair. The Independent variable of interest in each column is the standardized malaria positivity rate. Standardized malaria positivity rate is calculated over all sample years and districts. Malaria positivity rate data comes from the Tanzania Ministry of Health's National Portal (NHP). These data are malaria positivity rate among pregnant women during their first antenatal care (ANC) visit (Thawer et al., 2020). The dependent variable in all panels and columns, except Panel F, is the standardized exam score for the subject. The column header denotes the subject. The total score in the second column is the sum of exam scores in each subject that the student takes the exam. Standardized exam score for each subject is calculated over the sample of students in the corresponding exam-year and subject pair. Letter grades provided by the National Examination Council of Tanzania (NECTA) are converted into numeric points with letter grade "A" receiving four numeric points, letter grade "B" receiving three points, letter grade "C" receiving two points, letter grade "D" receiving one point, and letter grade "E" receiving zero point. Except third column, each specification in all columns includes student, exam-level, and exam-year fixed-effects in Panel A, Panel B, Panel C, and Panel D. In Panel A, Panel B, Panel C, and Panel D exam-level fixed-effects in third column are replaced with a fixed-effect specific to each subject and exam-level pair. The sample contains data from the Standard Four National Assessment (SFNA) exams conducted in 2015 and 2016 and the Primary School Leaving Examination (PSLE) exams conducted in 2018 and 2019. The candidates who appear for SFNA exams conducted in 2015 are longitudinally linked to PSLE exams conducted in 2018 and those who appear for SFNA exams conducted in 2016 are linked to PSLE exams conducted in 2019. Number of observations changes across the columns as not all students appear for all exams and for some students, NECTA does not report average letter grades. Panel A repeats estimates in Table 2. In Panel B, students appearing for PSLE exams are assigned the average malaria positivity rate in August and September of the exam-year in which they appear for exams. In Panel C, only those students who have a perfect string match of their names across SFNA and PSLE exams are part of the estimating sample. In Panel D, the malaria positivity rate in October of the exam-year in which the candidate appears for PSLE exams and the malaria positivity rate in December of the exam-year in which the candidate appears for SFNA exams are used. Panel E replaces student fixed-effects with student time-trends. Specification in Panel F has an indicator variable for whether the student is absent for the exam as the dependent variable.

Finally, Panel F of Table 3 presents an estimate for whether students who reside in districts with relatively higher levels of malaria infections fail to appear for exams. Specification in this panel is a variant of Equation 1 where the dependent variable is an indicator for whether the student is absent for the exam or not. The specification is estimated on the pooled sample and includes fixed-effects specific to each exam-level. The point estimate on the malaria positivity rate variable is not statistically significant. This result shows that the worsening of student performance on high-stakes exams at the primary school level is not influenced by the selection of students in terms of them not appearing for these exams.

To summarize, the results in this subsection establish a large and statistically significant worsening of performance on high-stakes primary school examinations when students are potentially exposed to higher levels of malaria infection in their proximate geography. This conclusion is robust to changing the analytical sample, changing the measure of potential malaria exposure, and estimating a more demanding empirical specification. I also do not find that the decline in academic achievement is driven by students selecting to appear for the exams. In the following subsection, I leverage detailed student-level data to uncover significant heterogeneity in the main effect of potential malaria exposure on student performance.

5.2 Heterogeneity Analysis

In this subsection, I present and discuss heterogeneity analysis. I begin by noting that the decline in student performance due to potential exposure to elevated levels of malaria infection is observed across almost all the subjects. These point estimates are presented in Table 1 and I also report them in Panel A of Table 4. As was noted above, the worsening of academic performance on high-stakes exams for primary school students is relatively more pronounced for reading (English) subjects than for quantitative (mathematics) subjects (p-value < 0.001). I do not find that the likelihood of failing the exams also increases when students' potential malaria exposure intensifies. The point estimate from estimating specification in Equation 1 with an indicator variable for whether the student failed the exam is reported in Panel B of Table 4.

On examining whether the estimated effect of potentially higher malaria infection exposure differs across male and female students, I observe that the effect of potential malaria exposure on academic performance does not differ by sex of the student consistently across different subjects. These point estimates are reported in Panel B of Table 4. While the worsening of exam performance due to potential exposure to higher levels of malaria infection is more pronounced for female students in mathematics examinations, the opposite is true for English examinations. The difference in total and average scores is also not statistically significantly different across students of different sexes.

Table 4: Heterogeneity Analysis

	Failed Exam	Total Score	All Subjects	Average Score	English	Kiswahili	Mathematics	Science	Social Knowledge
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Baseline									
Positivity Rate (SD)		-0.029***	-0.021***	-0.024***	-0.036***	-0.030***	-0.031***	0.016**	-0.028***
		(0.008)	(0.006)	(0.007)	(0.008)	(0.005)	(0.007)	(0.008)	(0.007)
Adj. R^2		0.521	0.495	0.472	0.426	0.482	0.458	0.321	0.337
N		2,332,647	15,240,211	2,299,700	2,310,316	2,310,416	2,310,321	2,310,424	2,310,166
Panel B: Heterogeneity by Sex									
Positivity Rate (SD)		-0.030***	-0.022***	-0.023***	-0.035***	-0.036***	-0.026***	0.016**	-0.029***
		(0.008)	(0.006)	(0.007)	(0.008)	(0.006)	(0.007)	(0.008)	(0.008)
$\mathbb{1}$ (Female)		0.001	0.001	-0.002	-0.002	0.012***	-0.010**	-0.001	0.002
×Positivity Rate (SD)		(0.005)	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)
Adj. R^2		0.521	0.495	0.472	0.426	0.482	0.458	0.321	0.337
N		2,332,647	15,240,211	2,299,700	2,310,316	2,310,416	2,310,321	2,310,424	2,310,166
Panel C: Heterogeneity by Regional GDP Per-capita									
Positivity Rate (SD)		0.003	0.008	0.008	0.005	-0.011	-0.004	0.042***	-0.009
		(0.009)	(0.008)	(0.009)	(0.010)	(0.007)	(0.008)	(0.009)	(0.009)
$\mathbb{1}$ ($\geq$ Median Real GDP)		-0.078***	-0.072***	-0.078***	-0.101***	-0.047***	-0.066***	-0.063***	-0.045***
×Positivity Rate (SD)		(0.015)	(0.013)	(0.015)	(0.017)	(0.011)	(0.014)	(0.015)	(0.015)
Adj. R^2		0.521	0.496	0.472	0.426	0.482	0.458	0.321	0.337
N		2,332,647	15,240,211	2,299,700	2,310,316	2,310,416	2,310,321	2,310,424	2,310,166
Panel D: Heterogeneity by Age									
Positivity Rate (SD)		-0.031***	-0.015**	-0.016**	-0.040***	-0.033***	-0.021***	0.034***	-0.030***
		(0.008)	(0.007)	(0.008)	(0.009)	(0.006)	(0.007)	(0.008)	(0.008)
$\mathbb{1}$ (PSLE Exam)		0.005	-0.018***	-0.023***	0.012*	0.009*	-0.027***	-0.050***	0.005
×Positivity Rate (SD)		(0.007)	(0.006)	(0.007)	(0.007)	(0.005)	(0.006)	(0.007)	(0.006)
Student FE		✓	✓	✓	✓	✓	✓	✓	✓
Exam-level FE		✓		✓	✓	✓	✓	✓	✓
Exam-year FE		✓	✓	✓	✓	✓	✓	✓	✓
Subject × Exam-level FE			✓						
Adj. R^2		0.521	0.496	0.472	0.426	0.482	0.458	0.322	0.337
N		2,332,647	15,240,211	2,299,700	2,310,316	2,310,416	2,310,321	2,310,424	2,310,166
Panel E: Failed exam									
Positivity Rate (SD)	-0.000								
	(0.001)								
Student FE	✓								
Exam-year FE	✓								
Subject × Exam-level FE	✓								
Adj. R^2	0.288								
N	15,389,426								

Notes: Heteroskedasticity robust standard errors clustered by the school (examination center) are in parentheses. (* p<.10 ** p<.05 *** p<.01). Each observation in all columns corresponds to a student-subject pair. The Independent variable of interest in each column is the standardized malaria positivity rate. Standardized malaria positivity rate is calculated over all sample years and districts. Malaria positivity rate data comes from the Tanzania Ministry of Health's National Portal (NHP). These data are malaria positivity rate among pregnant women during their first antenatal care (ANC) visit (Thawer et al., 2020). The dependent variable in all panels and columns, except Panel F, is the standardized exam score for the subject. The column header denotes the subject. The total score in the second column is the sum of exam scores in each subject that the student takes the exam. Standardized exam score for each subject is calculated over the sample of students in the corresponding exam-year and subject pair. Letter grades provided by the National Examination Council of Tanzania (NECTA) are converted into numeric points with letter grade "A" receiving four numeric points, letter grade "B" receiving three points, letter grade "C" receiving two points, letter grade "D" receiving one point, and letter grade "E" receiving zero point. Except third column, each specification in all columns includes student, exam-level, and exam-year fixed-effects in Panel A, Panel B, Panel C, and Panel D. In Panel A, Panel B, Panel C, and Panel D exam-level fixed-effects in third column are replaced with a fixed-effect specific to each subject and exam-level pair. The sample contains data from the Standard Four National Assessment (SFNA) exams conducted in 2015 and 2016 and the Primary School Leaving Examination (PSLE) exams conducted in 2018 and 2019. The candidates who appear for SFNA exams conducted in 2015 are longitudinally linked to PSLE exams conducted in 2018 and those who appear for SFNA exams conducted in 2016 are linked to PSLE exams conducted in 2019. Number of observations changes across the columns as not all students appear for all exams and for some students, NECTA does not report average letter grades. Panel A repeats estimates in Table 2. In Panel B, the specification in Equation 1 is modified to include an interaction between the sex of the student and malaria positivity rate among pregnant mothers in students' residence district. In Panel C, specification in Equation 1 is modified to include an interaction between the indicator for students' residence region having a real per-capita Gross Domestic Product (GDP) above the median and malaria positivity rate among pregnant mothers in students' residence district. Panel D replaces the independent variable of interest in Equation 1 to an interaction between the indicator for exam-level to PSLE and malaria positivity rate among pregnant mothers in students' residence district. In Panel E, the dependent variable in Equation 1 is replaced with an indicator for whether the student failed the exam.

Next, I examine if there is a widening of the academic performance gap between areas that are economically disadvantaged and those that are not. I use data from the Tanzania's National Bureau of Statistics (NBS). I draw upon data on regional real Gross Domestic Product (GDP) and population. Real GDP data comes from national account statistics and population data is drawn from the 2012 Housing and Population Census. For real per-capita regional GDP, I use data from 2015, the year immediately prior to the first year in my analytical sample. Using these data, I classify each region in the analytical sample as having below or above median real per-capita regional GDP. Point estimates in Panel C of Table 4 suggest that the effect of higher levels of potential malaria exposure is larger for economically better-off regions.

I present point estimates from heterogeneity analysis by age of the student in Panel D of Table 4. Students appear for PSLE exams after they successfully pass the SFNA exams. Examining if the point estimates differ by the exam-level allows me to study if the age of a student affects the impact that potential exposure to higher levels of malaria infection has on their academic achievement. Point estimates reflect that there is no consistent pattern of larger marginal effect by the age of the student. For some subjects, the marginal effects are more pronounced for younger students while the opposite is true for other subjects (column (7) of Panel D in Table 4 versus column(5) of Panel D in Table 4.)

Results in Panel D of Table 4 also point to the conclusion that malaria infection in Tanzania does not affect student performance only during the rainfall season. Tanzania has two climate regimes. The northern part of the country receives rainfall during two distinct periods – March to May (MAM) and October to December (OND). The rest of the country has an unimodal temporal rainfall distribution with a long rainfall season from November to April (NA). For the regions with the NA temporal rainfall distribution regime, the bulk of the rainfall is concentrated from December to February (DJF). I present regions with these distinct rainfall regimes in Figure 5. SFNA exams take place in November. This is during the rainfall season for regions both with unimodal and bimodal temporal rainfall distribution regimes. As I do not find a consistently more pronounced impact of potential exposure to higher levels of malaria infection for SFNA exams, the worsened student performance could not have been driven only by conditions where mosquitoes can breed easily.

I conclude this subsection by examining if the students are more likely to fail the exam when potentially exposed to higher levels of malaria infection in their residence district. I estimate a variant of specification in Equation 1 where I have replaced the dependent variable with an indicator for the student failing the exam. The point estimate suggests that there is no evidence that students are more likely to fail the exam when potentially exposed to higher levels of malaria infection. Taken together the results in this section suggest that there are significant heterogeneities in the estimated effect of potential exposure to higher levels of malaria infection on student performance. The main effect differs across students' sex, the economic performance of their residence region, and their age. In the following section, I study potential channels that may lead to the worsening of student performance when they are potentially exposed to

higher levels of malaria infection in the population geographically proximate to them.

5.3 Mechanisms Analysis

Results thus far show that potential exposure to higher levels of malaria infection in the days leading up to high-stakes primary school exams worsens student performance on these exams. I next turn to discuss estimates from specifications that aim to study mechanisms that may help explain the main effects. To study potential mechanisms, I use data from the National Panel Survey (NPS) of 2014/2015 conducted by the National Bureau of Statistics (NBS) of Tanzania. Descriptive statistics from NPS 2014 are presented in Table 6. Using these data, I estimate specification in Equation 2.

I present estimates from these specifications in Table 5. In the first column, the dependent variable is an indicator for whether the respondent missed school in the last two schooling weeks due to their own illness. I note that specifications in the first three columns are estimated conditional on the respondent reporting being in school. The point estimate in the first column whilst being positive is not statistically significant at conventional levels of statistical significance. In the second column, the dependent variable is an indicator for whether the respondent reports missing school due to the illness of another household member. The point estimate suggests that potential exposure to higher levels of malaria infection in students' geographically proximate population is positively and statistically significantly associated with a higher likelihood of them reporting missing school due to the illness of another household member.

A one standard deviation increase in the malaria positivity rate corresponds to an increase of approximately ten percentage points in the positivity rate. The point estimate in the second column reflects a more than doubling of respondents' likelihood over the sample mean of missing school due to the illness of another household member when the malaria infection rate increases by one standard deviation in their proximate geography. I do not find that increased potential exposure to malaria infection affects the likelihood of respondents reporting missing school due to either their own illness or the illness of another household member. Taken together, the results from these specifications suggest that the worsening of exam performance in the primary school high-stakes examinations could be driven by students missing school due to the illness of another household member on account of malaria exposure.

In the next column, the dependent variable is whether the respondent reported a visit to a healthcare provider in the last four weeks. Point estimate in this column suggest an increase in the likelihood of respondents reporting a visit to a healthcare provider in the last four weeks due to an increase in malaria infection in their residence district. Next, I examine how higher levels of malaria infection affect the likelihood of respondents reporting being hospitalized or having an overnight stay in the last 12 months from the date of the survey. While there is weak statistical evidence for higher levels of malaria infection leading to an increase in the likelihood of respondents being hospitalized or having an overnight stay, it

Table 5: Effect of Malaria on School Attendance and Health Outcomes

	Own Illness (1)	HH Member Illness (2)	Either Illness (3)	Healthcare Visit (4)	Hospitalized (5)	Hospitalized due to Malaria (6)
Positivity Rate (SD)	0.00169	0.00513	0.00681	0.03794	0.00536	0.02719
	(0.00383)	(0.00217)**	(0.00435)	(0.00790)***	(0.00314)*	(0.06894)
Weather Controls	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓
Household Controls	✓	✓	✓	✓	✓	✓
Dep. Var Mean	0.034	0.002	0.036	0.218	0.050	0.270
N	3,506	3,506	3,506	3,506	3,504	81

Notes: Heteroskedasticity robust standard errors are in parentheses. (* p<.10 ** p<.05 *** p<.01). Each observation is weighted by household sample weight. Each observation in all columns corresponds to a unique household member. Each column regresses the dependent variable on the malaria positivity rate measure, weather controls, individual controls, and household controls. Malaria positivity rate measure is standardized over all districts for the period during which the survey is conducted (January 2014 to September 2015). Dependent variables in the first three columns are conditional on the respondent being in school. The dependent variable in the last column is conditional on the respondent being hospitalized. The dependent variable in the first column is an indicator for whether the respondent missed school in the last two schooling weeks due to their own illness. The dependent variable in the second column is an indicator for whether the respondent missed school due to the illness of a household member. The dependent variable in the third column is an indicator for whether the respondent missed school due to either their own illness or the illness of another household member. The dependent variable in the fourth column is an indicator for whether the respondent visited a healthcare provider in the last four weeks. The dependent variable in the fifth column is an indicator for whether the respondent was hospitalized or had any overnight stay at a health facility in the last 12 months. The dependent variable in the sixth column is an indicator for whether the respondent was hospitalized or had an overnight stay at a health facility due to malaria. Weather controls contain cumulative precipitation and its square. Individual controls include an indicator for the respondent to be male and the age of the respondent. Household controls include the number of household members, an indicator for the household head being female, household expenditure on food during the last week (100,000 Tanzanian Shilling), an indicator for whether the household has access to electricity, household expenditure on education in the last 12 Months (100,000 Tanzanian Shilling), household asset index, and an indicator for whether the household owns a non-farm enterprise. I construct the household asset index as the sum of indicators for owning a bicycle, motorcycle, car/other vehicles (vans), tractor, computer, telephone, cellular, radio, television, refrigerator, and stove. Therefore, this index ranges from zero to eleven. Malaria positivity rate data comes from the Tanzania Ministry of Health's National Portal (NHP). These data are malaria positivity rate among pregnant women during their first antenatal care (ANC) visit (Thawer et al., 2020). Data on weather conditions comes from ERA5-Land (Hersbach et al., 2020). Data on other variables is drawn from the National Panel Survey (NPS) 2014/2015.

does not appear to be due to malaria.

Results in this section suggest a few things about the potential mechanisms that may lead to estimated effects. First, the worsening of exam performance may be driven by students missing school in the days leading up to exams as someone other than themselves in the household falls sick as malaria infection becomes more prevalent in the population. Second, malaria exposure may be accompanied by other ailments in addition to malaria infection, as evidenced by the increased likelihood of healthcare provider visits. Finally, the point estimates also suggest that the severity of illness, as seen through the increased likelihood of hospitalization or overnight stay in the hospital, may impose long-term penalties on academic achievement as students are more likely to miss school when hospitalized.

6 Discussion and Conclusion

Using data on student performance on high-stakes primary school exams, I study how malaria exposure in the days leading up to these exams affects student performance on these exams. In order to identify the causal effects of contemporaneous malaria exposure on student performance, I leverage within student variation in malaria exposure by longitudinally linking students across two primary school exams in Tanzania.

My results show that student performance worsens when they are potentially exposed to higher levels of malaria infection in the days preceding the exam days. Specifically, I find that a one standard deviation increase in students' residence district malaria positivity rate worsens student performance by 0.029 standard deviations. I find that this effect differs significantly by examination subject and by student demographics.

I also examine potential mechanisms that may help explain the main effects. Using survey data, I find that the worsening of student performance may be driven by students missing school due to the illness of another household member when the malaria infection in their geographically proximate population increases. I also show that the students are also more likely to visit a healthcare provider due to higher levels of malaria exposure. Increased hospitalization and overnight stays on account of higher levels of malaria exposure suggest that the worsening of performance may also have long-term consequences as students are forced to abstain from the exams and school.

While causally identifying the effect of contemporaneous malaria exposure on student achievement, my analysis leaves much to be desired. First, I do not have information on the actual malaria infection for students. Second, I do not observe how those who grade exams are impacted by malaria exposure. Finally, I do not observe student performance on different components of the exam. I hope future research is able to address these and other shortcomings.

References

Barat, Lawrence M, Nicole Whitehurst, Meera Venkatesan, Kim Connolly, Emmanuel Yamo, Paul Psychas, and Yves-Marie Bernard, "The U.S. President's Malaria Initiative's Support for Improving the Quality of Malaria Case Management Services: Fifteen Years of Progress and Learning," *Am J Trop Med Hyg*, November 2023.

Barofsky, Jeremy, Tobenna D. Anekwe, and Claire Chase, "Malaria eradication and economic outcomes in sub-Saharan Africa: Evidence from Uganda," *Journal of Health Economics*, 2015, *44*, 118–136.

Barreca, Alan I., "The Long-Term Economic Impact of In Utero and Postnatal Exposure to Malaria," *Journal of Human Resources*, 2010, *45* (4), 865–892.

Bleakley, Hoyt, "Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure," *American Economic Journal: Applied Economics*, April 2010, *2* (2), 1–45.

Cattaneo, Matias D., Richard K. Crump, Max H. Farrell, and Yingjie Feng, "Binscatter Regressions," 2023.

Cirera, Laia, Judit Vall Castelló, Joe Brew, Francisco Saúte, and Elisa Sicuri, "The impact of a malaria elimination initiative on school outcomes: Evidence from Southern Mozambique," *Economics & Human Biology*, 2022, *44*, 101100.

Cutler, David, Winnie Fung, Michael Kremer, Monica Singhal, and Tom Vogl, "Early-Life Malaria Exposure and Adult Outcomes: Evidence from Malaria Eradication in India," *American Economic Journal: Applied Economics*, April 2010, *2* (2), 72–94.

Donaldson, Dave and Adam Storeygard, "The View from Above: Applications of Satellite Data in Economics," *Journal of Economic Perspectives*, November 2016, *30* (4), 171–98.

Douglas, Nicholas M, Thomas R Burkot, and Ric N Price, "Malaria eradication revisited," *International Journal of Epidemiology*, 12 2021, *51* (2), 382–392.

Habyarimana, James, Ken Ochieng'Opalo, and Youdi Schipper, "The Contingent Electoral Impacts of Programmatic Policies: Evidence From Education Reforms in Tanzania," 2021.

Hersbach, Hans, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna De Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia de Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume, and Jean-Noël Thépaut, "The ERA5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, 2020, *146* (730), 1999–2049.

Hong, Sok Chul, "The Burden of Early Exposure to Malaria in the United States, 1850–1860: Malnutrition and Immune Disorders," *The Journal of Economic History*, 2007, *67* (4), 1001–1035.

Jain, Meha, "The Benefits and Pitfalls of Using Satellite Data for Causal Inference," *Review of Environmental Economics and Policy*, 2020, *14* (1), 157–169.

Klejnstrup, Ninja Ritter, Julie Buhl-Wiggers, Sam Jones, and John Rand, "Early life malaria exposure and academic performance," *PLOS ONE*, 06 2018, *13* (6), 1–16.

Kuecken, Maria, Josselin Thuilliez, and Marie-Anne Valfort, "Disease and Human Capital Accumulation: Evidence from the Roll Back Malaria Partnership in Africa," *The Economic Journal*, 12 2020, *131* (637), 2171–2202.

Lucas, Adrienne M., "Malaria Eradication and Educational Attainment: Evidence from Paraguay and Sri Lanka," *American Economic Journal: Applied Economics*, April 2010, *2* (2), 46–71.

Mboera, Leonard E. G., Emmanuel A. Makundi, and Andrew Y. Kitua, "Uncertainty in Malaria Control in Tanzania: Crossroads and Challenges for Future Interventions," *The American Journal of Tropical Medicine and Hygiene*, 2007, *77* (6 Suppl), 112 – 118.

Mordecai, Erin A, Sadie J Ryan, Jamie M Caldwell, Melisa M Shah, and A Desiree LaBeaud, "Climate change could shift disease burden from malaria to arboviruses in Africa," *The Lancet Planetary Health*, 2020, *4* (9), e416–e423.

Mtenga, Erica and Tejendra Pratap Singh, "Air Pollution and Student Achievement: Evidence from Tanzania," Aug 2022.

NECTA, "The Format For Standard Four National Assessment," https://necta.go.tz/exam_formarts/SFNA_FORMAT_english_online.pdf February 2018. Accessed: 2024-03-06.

__ , "Format For Primary School Leaving Examinations," https://necta.go.tz/exam_formarts/PSLE_ENG.pdf July 2020. Accessed: 2024-03-06.

Oesterholt, M. J. A. M., J. T. Bousema, O. K. Mwerinde, C. Harris, P. Lushino, A. Masokoto, H. Mwerinde, F. W. Mosha, and C. J. Drakeley, "Spatial and temporal variation in malaria transmission in a low endemicity area in northern Tanzania," *Malaria Journal*, Nov 2006, *5* (1), 98.

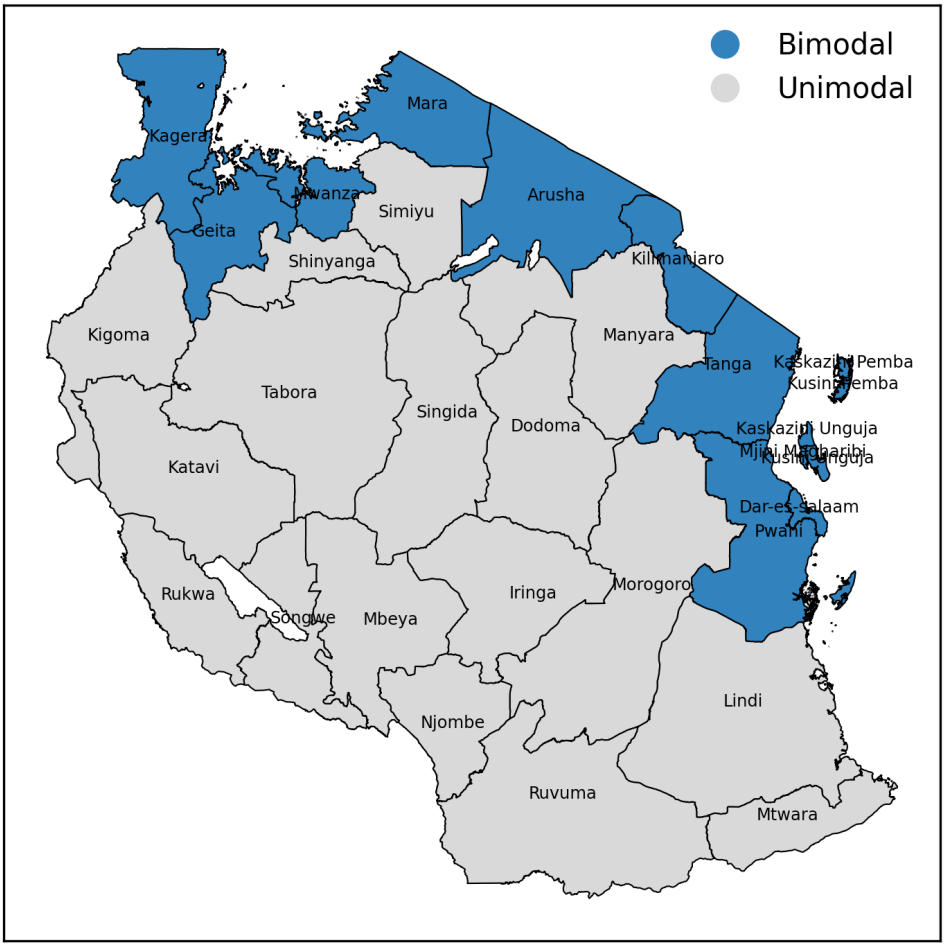
Rawlings, Samantha B., "Gender, Race, and Heterogeneous Effects of Epidemic Malaria on Human Capital and Income," *Economic Development and Cultural Change*, 2016, *64* (3), 509–543.

Thawer, Sumaiyya G., Frank Chacky, Manuela Runge, Erik Reaves, Renata Mandike, Samwel Lazaro, Sigsbert Mkude, Susan F. Rumisha, Claud Kumalija, Christian Lengeler, Ally Mohamed, Emilie Pothin, Robert W. Snow, and Fabrizio Molteni, "Sub-national stratification of malaria risk in mainland Tanzania: a simplified assembly of survey and routine data," *Malaria Journal*, May 2020, *19* (1), 177.

Venkataramani, Atheendar S., "Early life exposure to malaria and cognition in adulthood: Evidence from Mexico," *Journal of Health Economics*, 2012, *31* (5), 767–780.

Veras, Henrique, "Wrong place, wrong time: The long-run effects of in-utero exposure to malaria on educational attainment," *Economics & Human Biology*, 2022, *44*, 101092.

Figure 5: Rainfall Regimes in Tanzania



Note: This figure presents the spatial distribution of rainfall regimes in Tanzania. Bimodal temporal rainfall distribution regions receive rainfall during two distinct periods in the year – March to May (MAM) and October to December (OND). These regions are filled with navy color in the figure. The rest of the regions have an unimodal temporal rainfall distribution. Unimodal rainfall distribution regions receive rainfall from November to April (NA). These regions are filled with grey color in the figure. Shapefiles data comes from Global Administrative Areas (GADM).

Table 6: Summary Statistics NPS 2014

	N	Mean	SD	Min	Max
Outcomes					
Miss School in Last Two Weeks Due to Own Illness	4,461	0.038	0.190	0.00	1.00
Miss School in Last Two Weeks Due to Illness of a HH Member	4,461	0.003	0.053	0.00	1.00
Miss School in Last Two Weeks Due to Own Illness or Illness of a HH Member	4,461	0.040	0.197	0.00	1.00
Visited a health care provider in the last 4 weeks	16,245	0.222	0.415	0.00	1.00
Hospitalized or had an Overnight Stay	16,236	0.054	0.226	0.00	1.00
Hospitalization or Overnight Stay due to Malaria	805	0.294	0.456	0.00	1.00
Individual Controls					
Female	16,285	0.511	0.500	0.00	1.00
Age	16,285	21.912	18.842	0.00	100.00
Household Controls					
Number of Members in the Household	16,285	6.352	3.488	1.00	33.00
Female Household Head	16,285	0.240	0.427	0.00	1.00
Household Expenditure on Education in last 12 Months (100,000 Tanzanian Shilling)	15,794	6.925	51.110	0.00	2903.48
Household has Access to Electricity	16,285	0.326	0.469	0.00	1.00
Household Expenditure on Food Last One Week (100,000 Tanzanian Shilling)	16,167	0.289	0.259	0.00	2.21
Household Asset Index	16,285	2.743	1.740	0.00	11.00
Household Owns a Non-Farm Enterprise	16,285	0.500	0.500	0.00	1.00
Independent Variable					
Malaria Positivity Rate	13,126	9.357	8.603	0.10	44.00

Notes: Each observation is weighted by household sample weight. Each observation in all columns corresponds to a unique household member. Malaria positivity rate measure is standardized over all districts for the period during which the survey is conducted (January 2014 to September 2015). Variables in the first three rows of the outcomes panel are conditional on the respondent being in school. The variable in the last row of the outcomes panel is conditional on the respondent being hospitalized. The variable in the first row of the outcomes panel is an indicator for whether the respondent missed school in the last two schooling weeks due to their own illness. The variable in the second row of the outcomes panel is an indicator for whether the respondent missed school due to the illness of a household member. The variable in the third row of the outcomes panel is an indicator for whether the respondent missed school due to either their own illness or the illness of another household member. The variable in the fourth row of the outcomes panel is an indicator for whether the respondent visited a healthcare provider in the last four weeks. The variable in the fifth row of the outcomes panel is an indicator for whether the respondent was hospitalized or had any overnight stay at a health facility in the last 12 months. The variable in the sixth row of the outcomes panel is an indicator for whether the respondent was hospitalized or had an overnight stay at a health facility due to malaria. The individual controls panel includes an indicator for the respondent to be male and the age of the respondent. The household controls panel includes the number of household members, an indicator for the household head being female, household expenditure on food during the last week (100,000 Tanzanian Shilling), an indicator for whether the household has access to electricity, household expenditure on education in the last 12 Months (100,000 Tanzanian Shilling), household asset index, and an indicator for whether the household owns a non-farm enterprise. I construct the household asset index as the sum of indicators for owning a bicycle, motorcycle, car/other vehicles (vans), tractor, computer, telephone, cellular, radio, television, refrigerator, and stove. Therefore, this index ranges from zero to eleven. Malaria positivity rate data comes from the Tanzania Ministry of Health's National Portal (NHP). These data are malaria positivity rate among pregnant women during their first antenatal care (ANC) visit (Thawer et al., 2020). Data on other variables is drawn from the National Panel Survey (NPS) 2014/2015.