The Impact of Malaria Episodes and Treatment Regimens on Adverse Pregnancy Outcomes in Ugandan Women

Asmith Joseph

**Authors**

* Asmith Joseph (ORCID: 0009-0004-6875-0868)

**Author affiliations**

1. College of Public Health, University of Georgia, Athens, GA, USA.

Corresponding author: asmith.joseph@uga.edu

Disclaimer: The opinions expressed in this article are the author’s own and don’t reflect their employer.

# 1. Summary/Abstract

*Write a summary of your project.*

# 2. Introduction

## 2.1 General Background Information

Malaria is a life-threatening disease caused by Plasmodium parasites, which are transmitted to humans through the bite of an infected female Anopheles mosquito (Talapko et al., 2019). According to the World Health Organization, in 2022, approximately 249 million malaria cases were reported across 85 endemic countries—an increase of 5 million cases compared to 2021. Notably, Uganda alone contributed more than 597,000 cases during that period (WHO, 2023). Despite significant efforts in prevention and treatment, malaria remains a major global health challenge. Vulnerable groups such as pregnant women, children, and immunocompromised individuals are at the highest risk, and Africa continues to bear the heaviest burden of this disease. Pregnant women with malaria tend to experience more severe clinical symptoms and poorer outcomes, including heightened risks of miscarriage, fetal loss, premature birth, low birth weight in newborns, and neonatal death (Caroline Lin Lin Chua et al., 2021). A multifaceted approach incorporating insecticide-treated bed nets (ITNs) and chemoprevention strategies has proven effective in mitigating the adverse health outcomes associated with malaria during pregnancy. Despite this progress, a 2017 report reveals that only 22% of pregnant women in Sub-Saharan Africa received an entire course of IPTp (Bauserman et al., 2019), underscoring a significant disparity in treatment access. Addressing this issue, the study seeks to: (a) examine the association between the frequency of malaria episodes during pregnancy and adverse birth outcomes, such as preterm birth and stillbirth; (b) assess whether infants born to mothers treated with DP are less likely to experience low birth weight compared to those treated with SP; and (c) explore the relationship between placental malaria and adverse birth outcomes, including preterm birth and low birth weight.

## 2.2 Description of data and data source

The dataset originates from a double-blind, randomized controlled trial evaluating monthly intermittent preventive treatment of malaria (IPTp) using either sulfadoxine-pyrimethamine (SP) or dihydroartemisinin-piperaquine (DP) in HIV-uninfected pregnant women and their infants. Data were rigorously collected throughout the trial, with detailed records capturing maternal health, pregnancy outcomes, and infant details. In cases of multiple births, separate records were created for each infant during postpartum visits, ensuring comprehensive outcome documentation. The dataset comprises 782 observations and an extensive array of variables, enabling advanced modeling of the impact of malaria episodes and treatment regimens on adverse pregnancy outcomes.

The dataset encompasses several key variable categories. Maternal health and pregnancy outcomes are documented through unique participant identifiers, study arm assignment, and clinical indicators such as preeclampsia, various forms of postpartum hemorrhage (vaginal and cesarean), blood transfusions, laceration/episiotomy, congenital malformations, spontaneous abortions, stillbirths, and neonatal deaths, along with gestational parameters like the first day of the last menstrual period and estimated delivery dates. Reproductive history is captured via gravidity, parity, and counts of full-term births, preterm births, and abortions (both elective and spontaneous). Infant and child outcomes include details on delivery year, mode and location of delivery, birth complications, survival status, sex, and birth weight for up to seven children per participant. Malaria-specific measures record the total number of malaria episodes during pregnancy, infection rates per person-year, and results from routine blood smears, including the proportion of Plasmodium-positive. Additionally, the dataset contains information on maternal risk factors and comorbidities—such as education level, substance use, and various chronic conditions—and details on delivery complications and medications administered during labor. The data are publicly available through ClinEpiDB (Release #24, dated August 30, 2022).

## 2.3 Questions/Hypotheses to be addressed

The objective of this study is to examine the relationship between malaria exposure, treatment regimens, and adverse pregnancy outcomes in HIV-uninfected pregnant women receiving intermittent preventive treatment for malaria. Specifically, we will investigate whether an increased number of malaria episodes is associated with higher rates of preterm birth and stillbirth and whether the dihydroartemisinin-piperaquine (DP) regimen reduces these episodes compared to sulfadoxine-pyrimethamine (SP). We hypothesize that more malaria episodes correlate with an increased risk of adverse outcomes (Hypothesis 1) and that DP treatment results in fewer malaria episodes than SP (Hypothesis 2).

In addition, the study will explore infant health outcomes and the role of placental malaria. We will test whether infants born to DP-treated mothers are less likely to have low birth weight and whether neonatal mortality rates differ between the treatment groups. We hypothesize that DP treatment will improve infant outcomes by reducing the incidence of low birth weight (Hypothesis 3) and that neonatal mortality rates will vary between the SP and DP arms (Hypothesis 4). Finally, we will assess whether placental malaria is associated with adverse birth outcomes, such as preterm birth and low birth weight, and if its prevalence differs by treatment regimen, with the expectation that placental malaria increases the risk of these outcomes (Hypothesis 5).

# 3. Methods

*Describe your methods. That should describe the data, the cleaning processes, and the analysis approaches. You might want to provide a shorter description here and all the details in the supplement.*

## 3.1 Schematic of workflow

## 3.2 Data aquisition

The data for this study were obtained from ClinEpiDB, a public repository, specifically from Release #24 dated August 30, 2022. This dataset originates from a double-blind randomized controlled trial evaluating intermittent preventive treatment of malaria (IPTp) in HIV-uninfected pregnant women. The dataset, which includes 782 observations along with comprehensive maternal and infant health variables, was directly downloaded from the ClinEpiDB website without the need for authentication. Detailed metadata and documentation were also accessed via the platform, ensuring a clear understanding of variable definitions and study protocols.

## 3.3 Data import and cleaning

The dataset was imported into R directly from its source using rigorously tested and standardized data-loading functions. This approach ensured an accurate and complete transfer of information, providing you with confidence in the integrity of the data. We verified the dataset’s structure and consistency upon import, confirming the presence of 782 observations and cross-validating variable names with the corresponding metadata. To ensure proper formatting for subsequent analyses, date fields were converted into Date objects, and categorical variables were transformed into factors, preserving data integrity and preventing modeling errors.

Following the import and formatting, an initial inspection was conducted to assess the dataset’s structure and content. This review involved examining internal structures using functions such as str(), summary(), colnames(), and head(), as well as generating descriptive statistics to understand each variable’s central tendency, dispersion, and distribution. Visual inspections confirmed the data were correctly formatted and free from obvious entry errors, while exploratory visualizations identified potential outliers and anomalies. Recognizing the critical impact of missing data, we systematically quantified missing values by creating summary tables and using the missmap() function for visual assessment. Variables with more than 20% missing data were excluded to reduce potential bias. In contrast, those with less than 20% missing data were retained and appropriate imputation strategies were considered when necessary, ensuring that the final dataset was robust and reliable for subsequent statistical modeling.

## 3.4 Statistical analysis

# 4. Exploratory/Descriptive analyses

We conducted comprehensive exploratory and descriptive analyses to gain an initial understanding of the dataset’s structure, variable distributions, and data quality issues. In this phase, we summarized the demographic and clinical characteristics of the study population—including maternal age, parity, treatment allocation (SP vs. DP), and the distribution of malaria episodes during pregnancy—using descriptive statistics. Specifically, means, medians, standard deviations, and interquartile ranges were calculated for continuous variables, while frequencies and proportions were reported for categorical variables. Visualization techniques such as histograms, box plots, and bar charts helped identify outliers, skewed distributions, and other anomalies.

In addition, cross-tabulations were performed to explore preliminary associations between key exposure and outcome variables, such as the frequency of low birth weight across treatment groups. We also assessed missing data patterns, using this information to guide decisions on data imputation or excluding variables with excessive missingness. Together, these initial analyses provided critical insights that guided model selection and ensured the validity of subsequent inferential statistical tests.

Logistic and Count Regression Analyses (Maternal Health Outcomes)

This section examines the relationship between malaria episodes during pregnancy and adverse outcomes like preterm birth and stillbirth. The exposure variable will be the total number of malaria episodes during pregnancy, with preterm birth and stillbirth as binary outcomes. Logistic regression will assess these associations, adjusting for confounders such as maternal age, parity, and socioeconomic status. Another analysis will investigate whether treatment type (SP vs. DP) reduces malaria episodes. Here, the exposure variable will be treatment, and the outcome will be the total number of malaria episodes. Poisson or negative binomial regression will be used to compare episode counts across treatment groups, adjusting for factors like baseline malaria status, geographic region, and seasonal variations.

Chi-Square Tests and Logistic Regression (Infant Health Outcomes)

This section focuses on birth weight, gestational age, and neonatal mortality outcomes. First, we will assess whether infants born to mothers treated with DP are less likely to have low birth weight compared to those treated with SP. A chi-square test will initially compare proportions, followed by logistic regression to adjust for confounders such as maternal nutrition and gestational age. Next, neonatal mortality will be compared between treatment groups using chi-square tests for unadjusted differences, with logistic regression models controlling for factors such as birth complications, preterm birth, and birth weight. Lastly, we will examine whether malaria infection rates during pregnancy are associated with low birth weight and preterm birth. Logistic regression models will be used for this analysis, adjusting for maternal age, parity, and gestational age.

Logistic Regression and Chi-Square Tests (Malaria and Pregnancy Outcomes)

This section will assess whether placental malaria is associated with preterm birth and low birth weight. Logistic regression models will evaluate these associations while adjusting for relevant confounders, such as the number of malaria episodes during pregnancy and maternal health variables. Additionally, the difference in placental malaria prevalence between SP and DP treatment groups will be examined using chi-square tests for initial comparisons and logistic regression for adjustments, controlling for factors like the number of malaria episodes and regional differences.

# 5. Results

| **Variable** | **Dihydroartemisinin-piperaquine (DP)** N = 337*1* | **Sulphadoxine-pyrimethamine (SP)** N = 329*1* | **p-value***2* |
| --- | --- | --- | --- |
| **education\_level\_omrse** |  |  | 0.2 |
| A level | 5 (1.5%) | 4 (1.2%) |  |
| None | 21 (6.2%) | 35 (11%) |  |
| O level | 68 (20%) | 67 (20%) |  |
| Primary | 235 (70%) | 218 (66%) |  |
| Tertiary | 7 (2.1%) | 2 (0.6%) |  |
| University | 1 (0.3%) | 3 (0.9%) |  |
| **gravidity** |  |  |  |
| 1 | 73 (22%) | 84 (26%) |  |
| 2 | 90 (27%) | 70 (21%) |  |
| 3 | 39 (12%) | 57 (17%) |  |
| 4 | 46 (14%) | 37 (11%) |  |
| 5 | 40 (12%) | 36 (11%) |  |
| 6 | 23 (6.8%) | 17 (5.2%) |  |
| 7 | 16 (4.7%) | 13 (4.0%) |  |
| 8 | 8 (2.4%) | 11 (3.3%) |  |
| 9 | 2 (0.6%) | 4 (1.2%) |  |
| **parity** |  |  |  |
| 0 | 83 (25%) | 86 (26%) |  |
| 1 | 85 (25%) | 72 (22%) |  |
| 2 | 37 (11%) | 63 (19%) |  |
| 3 | 54 (16%) | 36 (11%) |  |
| 4 | 35 (10%) | 32 (9.7%) |  |
| 5 | 23 (6.8%) | 20 (6.1%) |  |
| 6 | 15 (4.5%) | 10 (3.0%) |  |
| 7 | 4 (1.2%) | 9 (2.7%) |  |
| 8 | 1 (0.3%) | 1 (0.3%) |  |
| **placental\_malaria** |  |  | <0.001 |
| No | 235 (70%) | 123 (37%) |  |
| Unknown | 10 (3.0%) | 12 (3.6%) |  |
| Yes | 92 (27%) | 194 (59%) |  |
| **preterm\_birth\_ontoneo** | 16 (4.7%) | 24 (7.3%) | 0.2 |
| **stillbirth** |  |  |  |
| No | 337 (100%) | 329 (100%) |  |
| **neonatal\_death** | 4 (1.2%) | 6 (1.8%) | 0.5 |
| **infant\_low\_birth\_weight\_2500\_g** | 24 (7.1%) | 26 (7.9%) | 0.7 |
| *1*n (%) | | | |
| *2*Fisher's exact test; Pearson's Chi-squared test | | | |

| **Characteristic** | **Dihydroartemisinin-piperaquine (DP)** N = 337*1* | **Sulphadoxine-pyrimethamine (SP)** N = 329*1* | **p-value***2* |
| --- | --- | --- | --- |
| total\_malaria\_episodes\_during\_pregnancy |  |  | < 0.001 |
| 0 | 323 (96%) | 257 (78%) |  |
| 1 | 14 (4.2%) | 57 (17%) |  |
| 2 | 0 (0%) | 14 (4.3%) |  |
| 3 | 0 (0%) | 1 (0.3%) |  |
| placental\_malaria |  |  | < 0.001 |
| No | 235 (70%) | 123 (37%) |  |
| Unknown | 10 (3.0%) | 12 (3.6%) |  |
| Yes | 92 (27%) | 194 (59%) |  |
| preterm\_birth\_ontoneo | 16 (4.7%) | 24 (7.3%) | 0.19 |
| infant\_low\_birth\_weight\_2500\_g | 24 (7.1%) | 26 (7.9%) | 0.77 |
| *1*n (%) | | | |
| *2*Fisher's exact test | | | |

## 5.1 Exploratory/Descriptive analysis

Note the loading of the data providing a **relative** path using the ../../ notation. (Two dots means a folder up). You never want to specify an **absolute** path like C:\ahandel\myproject\results\ because if you share this with someone, it won’t work for them since they don’t have that path. You can also use the here R package to create paths. See examples of that below. I generally recommend the here package.

## 5.2 Basic statistical analysis

*To get some further insight into your data, if reasonable you could compute simple statistics (e.g. simple models with 1 predictor) to look for associations between your outcome(s) and each individual predictor variable. Though note that unless you pre-specified the outcome and main exposure, any “p<0.05 means statistical significance” interpretation is not valid.*

## 5.3 Full analysis

*Use one or several suitable statistical/machine learning methods to analyze your data and to produce meaningful figures, tables, etc. This might again be code that is best placed in one or several separate R scripts that need to be well documented. You want the code to produce figures and data ready for display as tables, and save those. Then you load them here.*

# 6. Discussion

## 6.1 Summary and Interpretation

*Summarize what you did, what you found and what it means.*

## 6.2 Strengths and Limitations

*Discuss what you perceive as strengths and limitations of your analysis.*

## 6.3 Conclusions

*What are the main take-home messages?*

*Include citations in your Rmd file using bibtex, the list of references will automatically be placed at the end*

This paper (1) discusses types of analyses.

These papers (2,3) are good examples of papers published using a fully reproducible setup similar to the one shown in this template.

Note that this cited reference will show up at the end of the document, the reference formatting is determined by the CSL file specified in the YAML header. Many more style files for almost any journal [are available](https://www.zotero.org/styles). You also specify the location of your bibtex reference file in the YAML. You can call your reference file anything you like.

# 7. References

1. Leek JT, Peng RD. [Statistics. What is the question?](https://doi.org/10.1126/science.aaa6146) *Science (New York, N.Y.)*. 2015;347(6228):1314–1315.

2. McKay B, Ebell M, Billings WZ, et al. [Associations Between Relative Viral Load at Diagnosis and Influenza A Symptoms and Recovery.](https://doi.org/10.1093/ofid/ofaa494) *Open forum infectious diseases*. 2020;7(11):ofaa494.

3. McKay B, Ebell M, Dale AP, et al. [Virulence-mediated infectiousness and activity trade-offs and their impact on transmission potential of influenza patients.](https://doi.org/10.1098/rspb.2020.0496) *Proceedings. Biological sciences*. 2020;287(1927):20200496.