Latin American Data in Science, v.4, n.1
datainscience.com.br
ISSN: 2763-9290

Malaria notifications in Legal Amazon, Brazil (2003-2022): A Comprehensive Dataset for research & surveillance

Recebido: 02/07/2024 | Aceito: 06/10/2024 | Publicado: 13/11/2024
https://doi.org/10.53805/lads.v4i1.67

Kayo Henrique de C. Monteiro [1,2], Sebastião R. da S. Neto[1,2], Élisson da S. Rocha[1,2], Leonardo de C. Maia[3,4,5], Cássio Peterka[4], Vanderson S. Sampaio[6], Raphael Augusto Dourado[1,2], Theo Lynn[7], Patricia T. Endo[*1,2]

ABSTRACT

Malaria, a life-threatening disease, is preventable and curable. However, the World Health Organization reported alarming statistics for the year 2021, estimating 247 million cases of malaria worldwide, resulting in 619,000 deaths. In Brazil, cases of malaria are concentrated in the region of the Amazon rainforest, known as the Legal Amazon. In response to calls for more research to adapt malaria mitigation and eradication strategies to local conditions, this article presents a dataset based on data collected from the Malaria Epidemiological Surveillance Information System (SIVEP-Malaria) for the period from January 2003 to December 2022 on patients diagnosed with malaria. The pre-processed data set is composed of 1,251,309 records of confirmed malaria cases and included four attributes (Date, Municipality, Test results, and Notifications), attributes considered essential for the development and application of series modeling techniques temporal. By making this dataset available, it is intended to provide researchers with the necessary tools to develop effective strategies for mitigating and eradicating malaria in the region.

Keywords: Forecasting; Tropical diseases; Epidemiological data.

[1] Universidade de Pernambuco (UPE), Programa de Pós-Graduação em Engenharia da Computação (PPGEC), Recife, Brazil. patricia.endo@upe.br
[2] dotLAB Brazil, Caruaru, Brazil.
[3] Departamento de Doenças Transmissíveis (DEDT), Brazil.
[4] Secretaria de Vigilância em Saúde e Ambiente (SVSA), Brazil
[5] Ministério da Saúde (MS), Brazil.
[6] Instituto todos pela Saúde (ITpS), São Paulo, Brazil.
[7] Dublin City University, Dublin, Ireland.

DATA IMPORTANCE

- Development of Predictive Models: The dataset enables the creation of computational models that can accurately predict malaria cases and outbreaks. This is essential for anticipating and mitigating potential epidemics more effectively, enhancing the response of public health authorities.
- Understanding Transmission Dynamics: Researchers can use the dataset's rich information to deepen their understanding of malaria transmission dynamics. This includes analyzing geographical and temporal factors influencing disease incidence, contributing to more informed control and prevention strategies.
- Evaluation of Control Strategies: Analyzing the dataset helps evaluate the effectiveness of existing malaria control strategies, understand the impact of interventions, and identify potential areas for improvement. This is crucial for adjusting public health policies according to the specific needs of different regions.
- Identification of High-Risk Areas and Populations: Detailed data on malaria cases allow for the identification of high-risk areas and populations. This facilitates targeted interventions and more efficient resource allocation, ensuring that control and prevention measures effectively reach the most vulnerable communities.

MATERIALS AND METHODS

The data presented in this paper are sourced from the Malaria Epidemiological Surveillance Information System (SIVEP-Malaria) link: http://200.214.130.44/sivep_malaria/. SIVEP-Malaria acts as an epidemiological malaria surveillance system at the national level in all federal units of Brazil, its access is exclusive to professionals who work in the fight against epidemiological surveillance.

It comprises a comprehensive database that captures information on malaria cases, treatments and other pertinent factors, allowing complete and extensive monitoring of the disease across the country. Although SIVEP-Malaria has national coverage, it is important to note that more than 99% of malaria cases are concentrated in the Legal Amazon, which covers the states of Acre, Amapá, Amazonas, Mato Grosso, Maranhão, Pará, Rondônia, Roraima, and Tocantins. Due to its distinct environmental and geographic characteristics, the Amazon region serves as an ideal habitat for malaria-carrying mosquitoes, resulting in a higher incidence of malaria cases compared to other regions of Brazil.

Each record in the SIVEP-Malaria system consists of 47 attributes, Table 1 (in Anexs). The database encompasses the daily notifications of individuals diagnosed with malaria reported from January 2003 and December 2022, constituting a substantial volume of data comprising 44,736,283 records. Moreover, it allows for additional data aggregations, such as weekly, monthly, or yearly groupings, enabling a more focused analysis of malaria trends and patterns during the 2003-2022 period, as demonstrated in Figure 1 which illustrates the weekly notifications of confirmed malaria cases. The test result of each record is coded following the categories described in Table 2.

Focusing on the use of this data by time series models, we preprocessed the original dataset following the steps described in Figure 2. The first step involved consolidating the data from each year into a single dataset in order to streamline data analysis and manipulation. Next, only records pertaining to confirmed cases of malaria were selected for further inclusion. Subsequently, we discarded attributes that did not provide information about the municipality of the notification, date, or test result, since these attributes are not relevant for time series models. Finally, the variable NOTIFICATION DATE, initially in numeric format, was encoded into a standard DateTime format to facilitate subsequent analysis.

The final preprocessed dataset comprises 1,251,309 records of confirmed malaria cases, each record with four attributes: Date, Municipality, Test result (according to the codes shown in Table 2), and Notifications, as detailed in Table 3.

Figure 1.- Distribution of dataset records by year and malaria type.

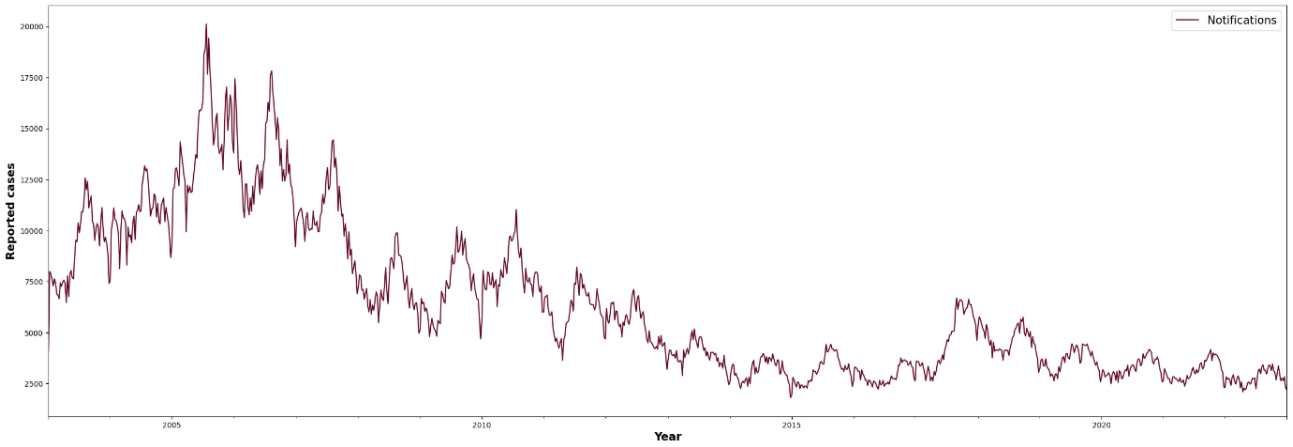
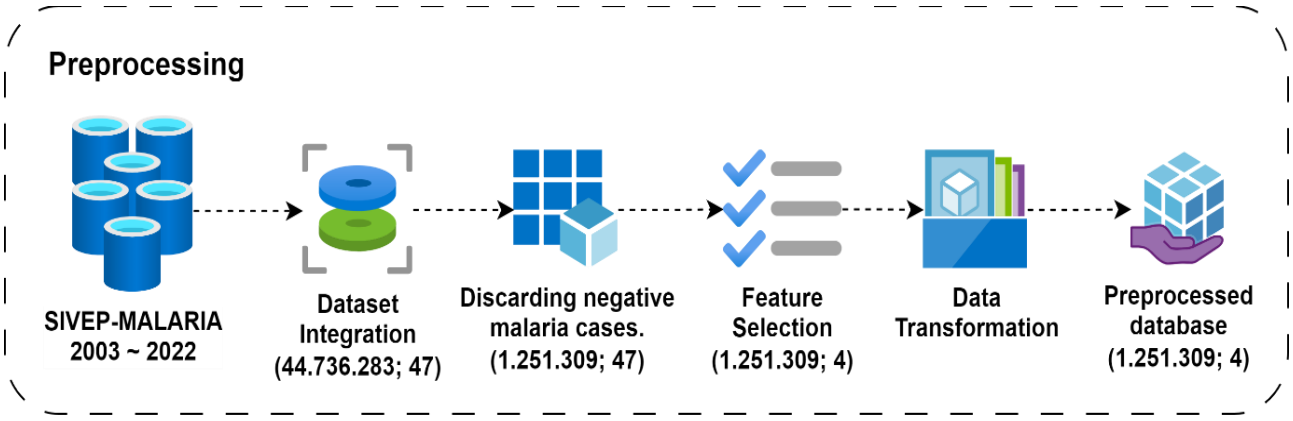


Table 2.- Attributes of the final preprocessed dataset.

Attribute	Description
DATE	Date of notification in the format yyyy/mm/dd.
MUNICIPALITY	Notifying municipality code (following IBGE codes).
TEST RESULT	Type of malaria according to the test result.
NOTIFICATIONS	Number of confirmed cases per day for each municipality and type of malaria.

Figure 2.- Preprocessing steps performed to build the final dataset.



The code used to merge and pre-process the dataset is publicly available on GitHub and can be accessed via the link: https://github.com/dotlab-brazil/Malaria_Legal_Amazon.

Table 3.- Test results of the original dataset.

Code	Results	Descriptions
01	-	Negative Cases
02	F	*Plasmodium falciparum*
03	F+FG	*Plasmodium falciparum + Plasmodium falciparum gametocytes*
04	V	*Plasmodium vivax*
05	F+V	*Plasmodium falciparum + Plasmodium vivax*
06	V+FG	*Plasmodium vivax + Plasmodium falciparum gametocytes*
07	FG	*Plasmodium falciparum gametocytes*
08	M	*Plasmodium malariae*
09	F+M	*Plasmodium falciparum + Plasmodium malariae*
10	O	*Plasmodium ovale*
11	No F	Not *Plasmodium falciparum*

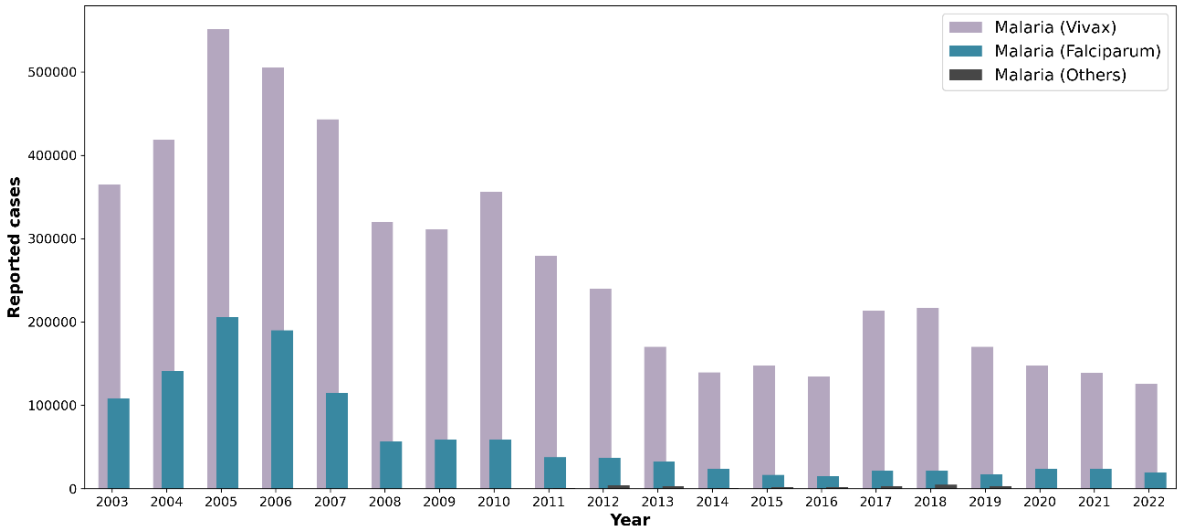
DATA DESCRIPTION

Daset

The preprocessed dataset, along with the corresponding data dictionary in English, are available on the Mendeley Data repository through the following link: https://data.mendeley.com/datasets/9n6b97fsbd/2

In Figure 3, we present the distribution of records in the dataset categorized by year and malaria type. The data spans from January 2003 to December 2022. It is important to note that cases were grouped based on test results (see Table 1). Specifically, we refer to cases with the results: F, F+FG, F+V, FG, and F+M as **Falciparum Malaria**; cases with the results V and V+FG as **Vivax Malaria**; and cases with the results M, O, and No F as **Other Malaria**. This categorization allows for a clearer understanding of the distribution and prevalence of different malaria types within the dataset.

Figure 3.- Distribution of dataset records by year and malaria type.



As shown in Figure 3, it can be observed that the predominant type of malaria during the whole time series is Plasmodium vivax. The highest number of confirmed cases was recorded in 2005, with a total of 757,514 cases. From 2010 to 2016, there was a decrease in the frequency of malaria cases. However, in 2017, there was a significant increase of 56.94% compared to the previous year, reaching a total of 237,955 cases. In 2018, there was a slight increase of 2.2% with 243,180 cases. This trend started to recede in 2019, with a substantial reduction of 21.83% reported malaria cases (190,093 in total). This reduction continued in the following years, with a reduction of 9.49% in 2020 (172,047 cases), 5% in 2021 (163,085 cases), and 11% in 2022 (145,547 cases).

While these consistent decreases in malaria cases indicate significant progress in malaria control efforts, it is important to acknowledge that the focus of health systems and resources shifted to address the challenges posed by the COVID-19 pandemic. This shift in priorities may have impacted the accuracy and reliability of malaria case reporting, potentially leading to an underestimation of the actual incidence of the disease. Therefore, caution should be exercised when interpreting the reported malaria cases during the pandemic period, taking into consideration the influence of the COVID-19 pandemic on surveillance and reporting mechanisms.

Figure 4 presents a visual representation of the number of malaria cases across the analyzed period, taking into account the location of the reporting municipalities. The data highlights the municipality of Manaus, in the state of Amazonas, as the location with the highest number of cases, totaling 601,700. Porto Velho, in the state of Rondônia, follows with 445,644 cases, and Cruzeiro do Sul, in the state of Acre, records 390,667 cases. These findings underscore the significant concentration of malaria cases in the Legal Amazon region, which spans across 716 municipalities in Brazil (IBGE, 2021).

Figure 4.- Incidence of malaria cases (all types) in the Legal Amazon between 2003 and 2022.

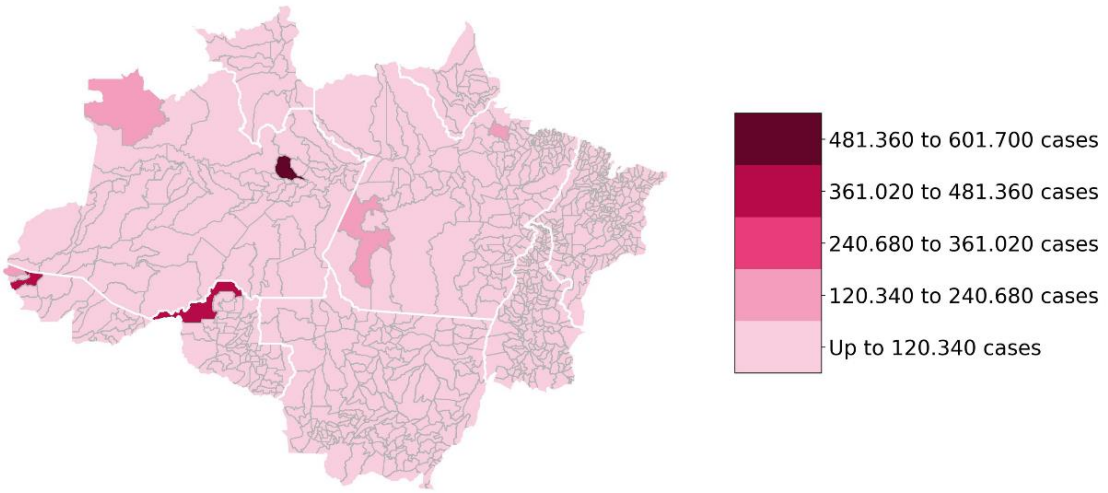
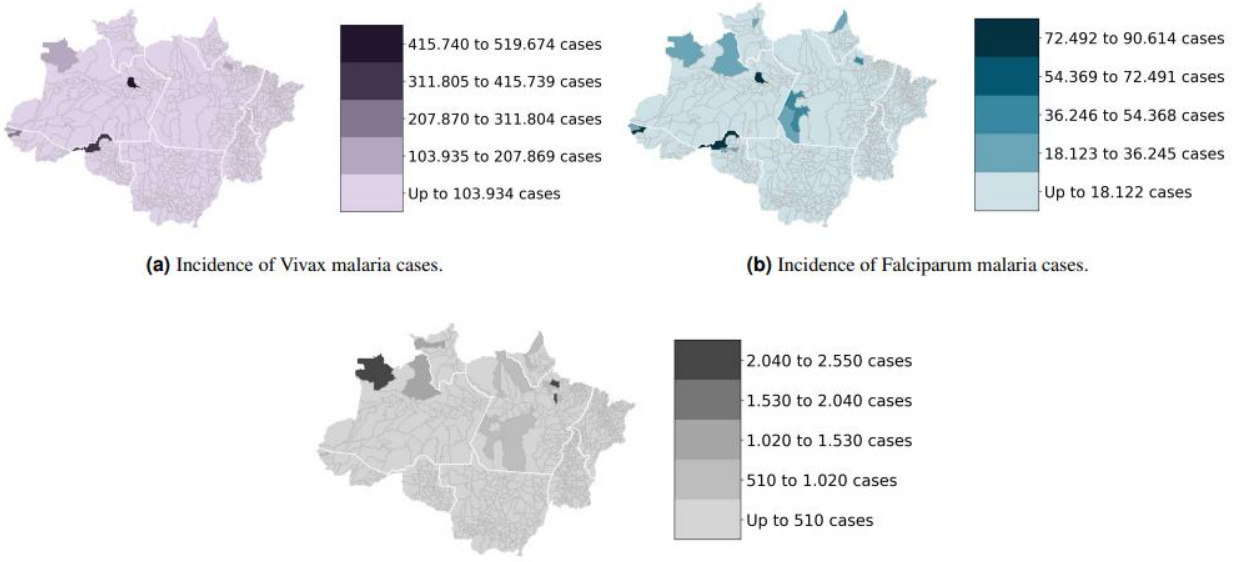


Figure 5 shows the incidence of malaria cases segmented by type in the Legal Amazon region. These maps provide valuable data on the geographical distribution of malaria cases, offering a more comprehensive understanding of the prevalence of different malaria types in each region.

Plasmodium vivax malaria, as shown in Figure 5.a was reported in 704 municipalities, with most cases concentrated in Manaus (Amazonas), Porto Velho (Rondônia), and Cruzeiro do Sul (Acre), in this order. Plasmodium falciparum malaria, as shown in Figure 5.b was reported in 601 municipalities, mainly in Cruzeiro do Sul, with the

highest number of cases, followed by Porto Velho and Manaus. To a lesser extent, other types of malaria (*Plasmodium malariae* and *Plasmodium ovale*) were reported in 295 municipalities, as shown in Figure 5.c. Notably, regions such as Oeiras do Pará and Anajás (state of Pará) and São Gabriel da Cachoeira (state of Amazonas) exhibit a higher incidence of these other types of malaria.

Figure 6.- Malaria cases in the Legal Amazon from 2003 to 2022, by municipality.



(a) Incidence of Vivax malaria cases.

(b) Incidence of Falciparum malaria cases.

(c) Incidence of others malaria types.

The Brazilian Amazon is comprised of regions considered endemic to malaria, with intense transmission rates (SVS, 2022). The high incidence of malaria cases in these regions is related primarily to favorable environmental conditions that promote mosquito proliferation and disease transmission cycles. Also, controlling malaria in these regions can present additional challenges, such as the presence of more virulent parasite strains and potential resistance to antimalarial drugs. Furthermore, human activities like deforestation and rural settlement can contribute to increased exposure and transmission of the disease.

It is important to note that the Amazon region encompasses a vast territorial expanse. This region is characterized by dense tropical forests and some densely populated urban areas. For instance, Manaus, the capital of Amazonas, has a high population density, 181 inhabitants/square kilometer in the year 2022 Instituto Brasileiro de Geografia e Estatística (IBGE, 2022).leading to a greater concentration of malaria cases in this area, similar to what occurs in Porto Velho, the capital of Rondônia, another area with considerable population density. On the other hand, the city of Cruzeiro do Sul has a lower population density compared to Manaus and Porto Velho but still reports a significant number of malaria cases.

These factors highlight the geographical heterogeneity of malaria distribution in Brazil, with varying prevalence rates depending on the specific parasite types involved. Consequently, it is imperative to implement effective strategies for malaria prevention, control, and treatment in the Legal Amazon region to reduce disease incidence and protect the health of the local population.

SUPPLEMENTARY MATERIALS

Dataset: https://data.mendeley.com/datasets/9n6b97fsbd/2

ACKNOWLEDGEMENTS

This work was partially funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE), and Universidade de Pernambuco (UPE), an entity of the Government of the State of Pernambuco focused on the promotion of teaching, research and extension.

REFERÊNCIAS

IBGE - Instituto Brasileiro de Geografia e Estatística. Amazônia legal. 2021. Available at: <https://www.ibge.gov.br/geociencias/cartas-e-mapas/mapas-regionais/15819-amazonia-legal.html?=&t=o-que-e>

IBGE - Instituto Brasileiro de Geografia e Estatística. Cidades: Manaus - AM. 2022. Available at: <https://cidades.ibge.gov.br/brasil/am/manaus/panorama>

SVS - Secretaria de Vigilância em Saúde. Panorama epidemiológico da malária em 2021: Buscando o caminho para a eliminação da malária no Brasil. Ministério da Saúde, Brazil. Ministério da Saúde, 2022. Available at: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/edicoes/2022/boletim-epidemiologico-vol-53-no17.pdf>

ANEXS

Table 1.- Attributes of the original database.

Attribute	Description
NOTIFICATION CODE	The notification number.
NOTIFICATION DATE	Date the test was performed and the patient received the diagnosis.
SEND DATE	The date the notification batch was entered into the national data file.
INPUTTED DATE	Date when the notification was entered into the system.
NOTIFICATION WEEK	Epidemiological week of the notification.
LAMINA TYPE	The type of detection.
NOTIFICATION STATE	The patient's state code according to the Instituto Brasileiro de Geografia e Estatística
NOTIFICATION MUNICIPALITY	The notifying municipality code, according to IBGE.
UNIT CODE	The code of the notifying health unit.
AGENT CODE	The code of the agent who carried out the test notification
DATE OF BIRTH	The patient's date of birth.
PATIENT AGE	Patient's age.
AGE SPECIFICATION	The format that the age is being specified. Which can be: D - Days M - Months Y - Years.
GENDER	The patient's gender.
RACE	The patient's Race / Color.
RESIDENCE COUNTRY	The country code refers to the patient's residential address according to IBGE.
RESIDENCE STATE	The state code refers to the patient's residential address according to IBGE
RESIDENCE MUNICIPALITY	The municipality code refers to the patient's residential address according to IBGE.
RESIDENCE LOCALITY	The code of the patient's place of residence.
SYMPTOMS	Whether the patient has symptoms or not.
SYMPTOMS DATE	The date on which the patient felt the first symptoms of malaria.
OCCUPATION CODE	Code of the main activity carried out by the patient in the last 15 days.
INFECTION STATE	The probable state of infection according to the IBGE code.
INFECTION COUNTRY	The country in which the patient was probably infected.
INFECTION MUNICIPALITY	It informs the municipality code referring to the probable municipality of infection acco
INFECTION LOCALITY	The Location code refers to the probable location of infection.
DATE TEST	Date the test was performed.
RESULT TEST	The result of the test.
PARASITE QUANTITY	The number of parasites per mm^3.
CROSS QUANTITY	The amount of parasitemia in crosses.
TREATMENT DATE	The treatment start date.
SCHEME	The code of the treatment scheme used.
HEMOPARASITES	The test result for other hemoparasites surveyed.
TEST	The type of test performed.
EXAMINER	The code of the professional who performed the test.
LVC IDENTIFIER	When the notification is an LVC case.

Attribute	Description
PREGNANT	If the patient is pregnant.
VIVAX	The patient received treatment for vivax malaria in the last 60 days before the notification
FALCIPARUM	The patient received treatment for falciparum malaria in the last 40 days before notification
EDUCATION LEVEL	The patient's education level.
WEIGHT	No information in the dictionary.
BLOD FORM	No information in the dictionary.
GAMETOCYTE TISSUE FORM	No information in the dictionary.
NU ACTIVITY G6PD	No information in the dictionary.
NU TOTAL HEMOGLOBIN	No information in the dictionary.
BREASTFEEDING	No information in the dictionary.
BREASTFEEDING TIME	No information in the dictionary.
DIMEA BREASTFEEDING	No information in the dictionary.